MC^2: MPEG-7 Content Modelling Communities

A thesis submitted for the degree of
Doctor of Philosophy

by

Damon Daylamani Zad
BSc MSc MBCS

School of Engineering and Design
Brunel University

May 2013

Abstract

The use of multimedia content on the web has grown significantly in recent years. Websites such as Facebook, YouTube and Flickr cater for enormous amounts of multimedia content uploaded by users. This vast amount of multimedia content requires comprehensive content modelling otherwise retrieving relevant content will be challenging. Modelling multimedia content can be an extremely time consuming task that may seem impossible particularly when undertaken by individual users. However, the advent of Web 2.0 and associated communities, such as YouTube and Flickr, has shown that users appear to be more willing to collaborate in order to take on enormous tasks such as multimedia content modelling. Harnessing the power of communities to achieve comprehensive content modelling is the primary focus of this research.

The aim of this thesis is to explore collaborative multimedia content modelling and in particular the effectiveness of existing multimedia content modelling tools, taking into account the key development challenges of existing collaborative content modelling research and the associated modelling tools. Four research objectives are pursued in order to achieve this; first, design a user experiment to study users' tagging behaviour with existing multimedia tagging tools and identify any relationships between such user behaviour; second, design and develop a framework for MPEG-7 content modelling communities based on the results of the experiment; third, implement an online service as a proof of concept of the framework; fourth, validate the framework through the online service during a repeat of the initial user experiment.

This research contributes first, a conceptual model of user behaviour visualised as a fuzzy cognitive map and, second, an MPEG-7 framework for multimedia content modelling communities (MC^2) and its proof of concept as an online service. The fuzzy cognitive model embodies relationships between user tagging behaviour and context and provides an understanding of user priorities in the description of content features and the relationships that exist between them. The MC^2 framework, developed based on the fuzzy cognitive model, is deep-rooted in user content modelling behaviour and content preferences. A proof of concept of the MC^2 framework is implemented as an online

service in which all metadata is modelled using MPEG-7. The online service is validated, first, empirically with the same group of users and through the same experiment that led to the development of the fuzzy cognitive model and, second, functionally against the folksonomy and MPEG-7 content modelling tools used in the initial experiment. The validation demonstrates that MC^2 has the advantages without the shortcomings of existing multimedia tagging tools by harnessing the ease of use of folksonomy tools while producing comprehensive structured metadata.

Acknowledgements

The research work in this doctoral thesis was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant “MC^2: MPEG-7 Content Modelling Communities” (EP/E034578/1) on which I was the sole Research Fellow. First and foremost, I would like to express my deepest gratitude and appreciation to *Professor Marios C. Angelides*, my supervisor and the Co-Investigator on the EPSRC project, for always being available anytime I needed his guidance, support and motivation throughout my doctoral studies. It has been an honour and a privilege to work under his supervision. I am also very grateful to *Dr Harry Agius,* my second supervisor and the Principal Investigator on the EPSRC project, for his unreserved support, encouragement and wisdom, without which this research may not have reached its goals. I would also like to extend my gratitude to *Cambridge Imaging Systems* for their valuable and generous co-operation during the project, to the users who gave a great deal of their time to take part in the experiments, to all the users who continue to help MC^2’s user base grow, and to my friends for their moral support and help whenever I needed it. I would like to dedicate this thesis to my *family* for their unconditional love, support and encouragement throughout my life and for always believing in me. A very special thank you goes to *Maryam* for her unconditional support and understanding.

List of Publications

Journal Articles

AGIUS, H., ANGELIDES, M.C. and **DAYLAMANI ZAD, D.** 2013. MC^2: A Framework and Service for MPEG-7 Content-Modelling Communities. *The Computer Journal*, 56(5), 593-616.

AGIUS, H., ANGELIDES, M.C. and **DAYLAMANI ZAD, D.** 2013. Experimenting with tagging and context for collaborative MPEG-7 metadata. *Multimedia Tools and Applications,* 62(1)**,** 143-177.

DAYLAMANI ZAD, D., ANGELIDES, M.C. and AGIUS, H. 2012. Personalise your Massively Multiplayer Online Game (MMOG) with Artemis. *ACM/Springer Multimedia Systems Journal,* 18(1)**,** 69-94.

Conference Proceedings

DAYLAMANI ZAD, D. 2007. Semantic Content Filtering using Self-Organizing Neural Networks. In: MYLONAS, P.P., WALLACE, M. and ANGELIDES, M.C. (Eds.) *IEEE-CS Proceedings of the 2nd International Workshop on Semantic Media Adaptation and Personalisation (SMAP 2007)*, Uxbridge, UK, December. IEEE-CS: Los Alamitos, California, USA, 253-256.

Book Chapters

DAYLAMANI ZAD, D., ANGELIDES, M.C. and AGIUS, H. (Forthcoming). Collaboration through gaming. In: ANGELIDES, M.C. and AGIUS, H. (Eds.) Handbook of Digital Games. Wiley-IEEE Press: Hoboken, New Jersey, USA.

DAYLAMANI ZAD, D. and AGIUS, H. 2011. An MPEG-7 Profile for Collaborative Multimedia Annotation. In: ANGELIDES, M.C. and AGIUS, H. (Eds.) *The Handbook of MPEG Applications: Standards in Practice.* Wiley: Chichester, West Sussex, UK, 263-291.

DAYLAMANI ZAD, D. and AGIUS, H. 2009. Collaborative movie annotation. In: FURHT, B. (Ed.) *Handbook of Multimedia for Digital Entertainment and Arts.* Springer: New York, New York, USA, 265-288.

DAYLAMANI ZAD, D. and AGIUS, H. 2009. Multimedia Metadata 2.0: Challenges of collaborative content modelling. In: ANGELIDES, M.C., WALLACE, M. and MYLONAS, P. (Eds.) *Advances in Semantic Media Adaptation and Personalization.* CRC Press: Boca Raton, Florida, USA, 1-20.

Work in Progress

DAYLAMANI ZAD, D., ANGELIDES, M.C. and AGIUS, H. Collaboration through multi-platform gaming with Artemis

Table of Contents

List of Figures

List of Tables

Chapter 1: Collaborative Multimedia Content Modelling

This chapter begins by examining the effectiveness of existing approaches for multimedia tagging tools and categorises them based on tagging (structured or unstructured) and context (individual or collaborative). The chapter also unravels future challenges with collaborative content modelling in three categories, namely, model stability, plasticity and accuracy. The chapter then draws its research aim and objectives and the research methods used in pursuing each objective.

1.1 Background

The use of multimedia on the Web has grown massively in recent years and Internet traffic is predicted to quadruple in the near future predominantly due to online video (Cisco, 2012). Video-on-demand Websites such as BBC iPlayer, 4oD and Sky Anytime+ in the UK and Hulu and Netflix in North America, along with online global video Websites such as YouTube and social networking Websites such as Facebook, have changed user habits and attitudes towards Internet video usage and the statistics are significant. While Google processes around 700,000 search results every minute (Fach, 2011), 60 hours of video are uploaded to YouTube every minute with 4 billion videos (100 million hours of video) viewed per day by over 27 million visitors (Grossman, 2012; YouTube, 2012). In one day, over 250 million photos are also uploaded to Facebook and 2.7 billion Likes and Comments are posted by 483 million daily active users. In total, Facebook stores over 100 petabytes of photos and videos (Ebersman, 2012).

Consequently, Web 2.0 can be said to have given rise to increasing use of multimedia content on the Web, leading to the emergence of many collaborative multimedia environments such as Flickr, YouTube and Del.icio.us where users are able to share and view multimedia content in a community. These folksonomic communities or social-tagging applications allow users to describe multimedia content using simple, flat metadata such as titles, categories and tags. These applications suggest that users are willing to model multimedia content in a collaborative manner and therefore it is possible to extend this approach to more advanced media content modelling applications, such as those based on the comprehensive, extensible MPEG-7 standard. In this way, the power of Web communities would be harnessed to overcome what is otherwise an intensely time consuming task for single users of creating, updating and maintaining content models for multimedia resources. When it is considered that non-expert users tend to add only 3-4 tags per content item on average (Jackson and Smith, 2012), it also becomes apparent that there is great potential for improving the richness of the content models through collaboration.

However, there is a semantic gap between the current folksonomic communities and real world semantics which stops the models created by these tools from being fully and comprehensively mapped. For example in YouTube, the user has no means of clearly defining temporal relationships or defining the time an event starts and finishes. Other challenges include overcoming spelling mistakes, different terms being used by different users to describe the same content, and creating community feeling so that users productively engage, which currently impose barriers for media retrieval and affect the usefulness of the content models. When these challenges are overcome, more detailed and accurate content models become possible and a greater amount of multimedia resources are able to be content modelled.

This steep increase in the volume and heterogeneity of multimedia content has necessitated their tagging to create metadata that describes the content thereby assisting the retrieval process (Edvardsen et al., 2009). Metadata now plays a critical role in the functionality of most digital information repositories (Mayernik et al., 2011; Ojokoh et al., 2011; Haslhofer and Klas, 2010; Rodriguez et al., 2009; Saathoff and Scherp, 2010). Each tag maps to one or more specific features of the content and metadata typically include the semantics of the content such as its substance, recording location, content type, and so on. In *unstructured metadata*, the tags consist of freely-entered textual descriptions, such as those associated with YouTube[1] videos, whereas in *structured*

[1] http://www.youtube.com

metadata, the tags are organised and related within a well-defined scheme, such as those prescribed by the Dublin Core[2] and MPEG-7 (ISO/IEC, 2003; ISO/IEC, 2004; ISO/IEC, 2005a) standards which try to enforce 'good practice' in metadata creation. In such standards, both content metadata and user metadata are typically prescribed (Laborie et al., 2010). Content metadata normally includes *low level* metadata (e.g. colour, texture, shape, melody), *high level* metadata (e.g. terms, concepts, keywords), and s*tructure* metadata (e.g. spatial and temporal segmentation, audio and video streams), while user metadata normally includes *user interactions* (e.g. user preferences, usage history) and *user context* (e.g. terminal, network, quality of service). The metadata scheme and its information will depend on the content type.

In contrast to the traditional subject indexing and cataloguing that takes place in libraries and museums, metadata is now generated not only by experts but also by creators and consumers of the content (Voss, 2007). Consequently, tagging of content features may be achieved either *individually*, that is, by a single user working alone, or *collaboratively*, that is by multiple users working together, typically within a community. The rise of Web 2.0 and services based on wikis, which allow the pages of a website to be modified by anyone at any time, have proven that global communities of users are not only able to work together effectively to create detailed, useful content, even minutiae, for the benefit of others, but do so voluntarily and without solicitation. Individual tagging is a time consuming process for a single user to undertake; therefore, multiple users can greatly reduce effort while also increasing the detail, quality and volume of the metadata. For example, in Flickr[3], users may tag their own and each other's photos and interact with the users in the community.

The dynamics of collaborative tagging are notably different to those that may be found in an individual tagging environment. Collaborative tagging enables some users to tag some resources or a part of a resource while other users are tagging other parts or different resources. In addition, some users can focus on a single content feature or a specialist subset of content features throughout the

---

[2] http://www.dublincore.org/
[3] http://www.flickr.com

resource repository while others can vary where they contribute across multiple content features as they see fit. These dynamics also enable groups of users within the community to work on similar content features that interest them (Tang et al., 2008). Collaboration also supports the fact that different users will work at different times depending on their geographical location, their habits and their engagements. In addition, the interactions between the users in a community may spark new ideas and points of views on details that might have been missed by one individual (Jung, 2008), thus collaboration also allows for corrections of tagging mistakes (Voss, 2007; Lee and Yong, 2007). In short, in a collaborative community, the users are diverse in terms of their motivation, interest, availability, experience, skill and knowledge, which enables the efficient use of their combined characteristics to cover the widest possible range of multimedia tagging.

This research explored collaborative modelling of multimedia content and in particular how effective existing tools are. After consideration of both the shortcomings of existing tools and future challenges faced with collaborative modelling, it undertook an experiment with over 50 users in order to examine how they work with existing multimedia tagging tools, both structured and unstructured, so that a conceptual model of user behaviour may be constructed that shows the relationships between user behaviour. Then the conceptual model of user behaviour was used to design, develop and implement the MC^2 (MPEG-7 Content Modelling Communities) framework and online service that caters for MPEG-7 content modelling communities. The remainder of this chapter presents a review of multimedia tagging tools, followed by a discussion of the challenges faced by collaborative content modelling, a presentation of the research aim and objectives, the research methods used for pursing each objective and a summary of the thesis.

1.2 Approaches for multimedia tagging tools

Considering the permutation of the tagging (structured or unstructured) with its context (individual or collaborative) enables a categorisation of multimedia tagging within four areas. Figure 1.1

illustrates the four areas along with example multimedia tagging tools for each category that embody the respective approach. Each category is now considered in turn.

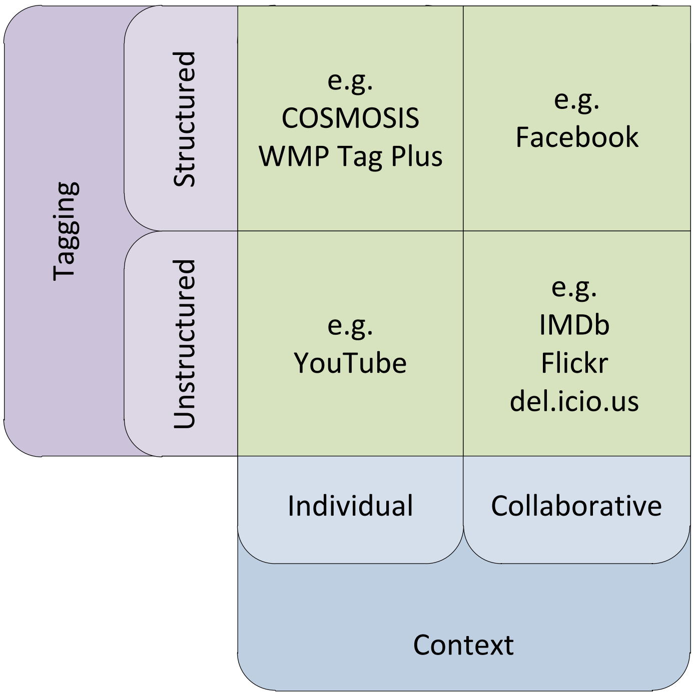


Figure 1.1: Tagging and Context in multimedia tagging tools

1.2.1 Individual-unstructured tagging

In this approach, the metadata is created and maintained by one user and no other users contribute to the metadata. The systems may be stand-alone systems, such as desktop applications, or Web-based. The metadata itself is unstructured, typically just a collection of keywords, without any categorisation or organisation. For example, YouTube provides a community for sharing video content on the Web and enables users to upload their videos, set age ratings, and enter descriptions and keywords. However, despite the presence of an online community with commenting capabilities, only the owner (uploader) of the video can add tags to it and users may not contribute to each other's tags. Furthermore, the nature of free text is a source of problems, i.e. the unstructured format, the ambiguous word usage, the cut-and-paste additions, the abbreviations, the inserted HTML/XML tags, the multimedia content, and the domain-specific terminology (Johnson et al., 2012). Extensive research has been conducted on automatic tagging of images (Carneiro and Vasconcelos, 2005; Celebi and Alpkoca, 2005; Viitaniemi and Laaksonen, 2007; Hu and Lam, 2013), videos (Tao and Embley, 2009; Wu and Li, 2011; Lu and Li, 2008; Tonelli et al., 2013) and biomedical multimedia resources (Névéol et al., 2011). The predominant method in these approaches is to recognise and extract relevant content features using various object and pattern recognition algorithms, then apply different textual and visual similarity matching algorithms with already tagged resources and propagate the related metadata to the new extracted features. Wu and Li (2011) combined visual features extracted from the visual track of video and keywords extracted from speech transcripts of the audio track to enhance the semantics of the generated metadata. Shin et al. (2010) followed an emotion prediction approach for automatic tagging of videos where colour and pattern are used to predict the emotional semantics associated with an image. Toti et al. (2012) describe a methodology for identifying characterising terms from a source text and automatically building an ontology around them, with the purpose of semantically categorising and clustering with similar documents through ontology alignment. These features are extracted using a colour quantisation and a multi-level wavelet transform, respectively. While automated techniques work much more quickly than manual tagging, they are not always as efficient. These methods (Hyun-seok et al., 2012; Fries et al., 2012; Daltayanni et al., 2012) rely on already tagged resources that have been tagged manually and therefore still need human involvement to initiate them; hence, they are semi-automatic at best. Moreover, if the feature extracted does not match already tagged features, then the tagging of it will not be accurate (Haq et al., 2012; Fei et al., 2012). This applies to all other features as well since the metadata generated through these systems is general and tends to be imprecise. Automated methods will recognise an object in a scene but they are generally not able to identify what this object is in relation to the semantics of the resource.

1.2.2 Individual-structured tagging

This approach is similar to the previous approach in that the tags are added and modified within the metadata by a single user; however, the metadata is structured and typically obeys a standardised scheme (Paschalakis et al., 2012; Ioakim et al., 2012; Bastan et al., 2012; Vretos et al., 2012; Lei, 2012; Yong-Hwan et al., 2012; Yoon, 2013; Waltl et al., 2013; Han et al., 2013). These systems are stand-alone systems, typically desktop applications, and are often not Web-based. For example, WMP Tag Plus[4], a Windows Media Player plug-in that adds support for reading and writing tags of MPEG-4, FLAC[5], Vorbis[6] and WavPack[7] files, is one such example. Similarly, COSMOSIS (Agius and Angelides, 2006) is a system based around MPEG-7 which supports extensive individual tagging. With this system, users can tag audio-visual (AV) content and extensively define the semantics of their content such as objects, events, temporal relations and spatial relations. Through separating semantic video content aspects via tabbed panes, COSMOSIS enables relevant information from each aspect of metadata to be displayed simultaneously while also permitting comfortable cross-referencing between sections. Because individual tagging is an intensely time consuming task for single users, involving creating, updating and maintaining large amounts of metadata for multimedia resources, many approaches have been proposed for automating some or all of this process. Ryu et al. (2002) presented an individual tagging system which accommodates an integrated metadata authoring environment based on MPEG-7 description schemes. It is equipped with an automatic content analysis module which allows for automatic temporal structuring of video at the shot level and semi-automatic hierarchical structuring at the higher levels.

1.2.3 Collaborative-unstructured tagging

In this approach, the metadata is created by multiple users working together through an online web-based system, although the metadata itself is unstructured. Web 2.0 has brought with it an increasing use of collaboratively-generated data which demonstrates that tasks such as multimedia tagging can also be tackled collaboratively such that a community creates, updates and maintains

[4] http://bmproductions.fixnum.org/wmptagplus/
[5] http://flac.sourceforge.net/
[6] http://www.vorbis.com/
[7] http://www.wavpack.com/

the metadata of multimedia resources. IMDb[8] is one example in this area. It is an online database of information relating to movies, television shows, actors, production crew personnel, video games and fictional characters featured in visual entertainment media. IMDb allows users to contribute and modify existing metadata, but the metadata is not structured other than being placed into various sections of an entry profile. All contributed metadata is vetted by IMDb administrators, but it does not show who contributed the metadata and there is no communication between users with regards the tagging. There are many categorisations of collaborative tagging within the research literature (Li and Lu, 2008), but *folksonomies* are perhaps the most well-known example of collaborative-unstructured tagging (Santos et al., 2013; Solskinnsbakk et al., 2012; Uddin et al., 2013; Vandic et al., 2012; Carmel et al., 2012; Lee et al., 2012; Semeraro et al., 2012; García et al., 2012; Rodenhausen et al., 2012). Folksonomies are user-generated taxonomies where tags are collaboratively created and managed by a broad cross-section of users including experts, creators and consumers of the content, to describe and categorise that content. Usually, freely chosen keywords are used instead of a controlled vocabulary (Voss, 2007). Consequently, the tags are highly individualistic and different users use a variety of terms to describe the same concept (Lee and Yong, 2007). del.icio.us[9] was one popular example of a folksonomy which enabled users to bookmark and tag pages they come across while browsing the Internet and to share these with other users. They could also add comments and descriptions to their tagged pages. AV content is now commonly found on various websites, e.g. blogs, and therefore del.icio.us provides a means to tag and add descriptions of the AV content. Another example is Flickr, which provides a means for photo enthusiasts at all levels to share their photos in original quality. Tools are provided for a user to tag their own photos and allow other users to tag their photos. Users may post comments on photos as well as demarcating areas of images with comments. Flickr also enables users to perform simple manipulations on their photos such as rotating and cropping. Zhang et al. (2009) propose a video blog management model which is

---

[8] http://www.imdb.com/

[9] At the time of writing, del.icio.us is in the process of being revamped as Delicious, accessed via http://www.delicious.com/, and it is unsure how much of the original functionality will remain.

comprised of automatic video blog tagging and user-oriented video blog search, extracting keywords from both the target video blog itself and relevant external resources. Others have worked on using mobile phones as the platform for folksonomy approaches. Zonetag (Naaman and Nair, 2008) is a prototype mobile application that uploads camera phone photos to Flickr and assists users with context-based tag suggestions derived from the community tagging activity on Flickr, based on the user's own tagging history and the tags associated with the location of the user. Another method in the folksonomy approach is proposed by Golder and Huberman (2006b), where they use a stochastic urn model, originally used to represent how diseases spread and contaminate, to model the formation and stabilisation of folksonomies. Their analysis illustrates the key factors involved in stabilising tag usage and forming a folksonomy. Finally, Begelman et al. (2006) investigate different factors in the formation of folksonomy and tag use by studying frequency of use and co-word clustering. They show that tagging actually mimics the behaviours observed in conventional classification systems. Social media approaches attempt to capitalise on social networks such as Facebook[10] by incorporating knowledge of the social network into the collaborative tagging process to improve the understanding of tag behaviours (Li and Lu, 2008). For example, Mika (2007) builds various ontologies from tags on the basis of concepts and communities. Ohmukai et al. (2005) presents a community-based ontology using both the metadata generated by users and their personal social networks. Pea et al. (2006) propose the DIVER Software Environment for Video Collaboration. Collaborative video analysis is supported through a web-enabled DIVER that allows for distributed access and annotation of digital video records from consumer digital cameras. The central work product in DIVER is known as a 'dive', which consists of a set of XML metadata pointers to segments of digital video stored in a database and their affiliated text annotations. William (2006) presents a collection-oriented metadata framework which is based on group and social networking effects for handling digital images. Yamamoto et al. (2008) present an approach for video scene tagging based on social activities associated with the content of video clips on the web using data

[10] http://www.facebook.com

from forums, user comments and blogs. Other studies have begun to incorporate the tag and entity relationships, including displaying related entities as hints to assist the user in finding the appropriate tags as those tags are being entered. Aurnhammer et al. (2006) introduce an approach that combines social tagging and visual features to extend the navigation possibilities of image archives. Yang et al. (2008) apply the shapers (Editors in Wikipedia) methodology in collaboration and negotiation applications, and find that shapers can greatly assist a group in forming consensuses during negotiations and collaborations. Kwon (2009) proposes a methodology that increases user performance in terms of costs associated with building consensus and successful negotiation rates. Maleewong, Anutariya and Wuwongse (2008) present a collective intelligence approach to collaborative knowledge creation which enables multiple users to collaboratively create knowledge in a systematic and dynamic process.

1.2.4 Collaborative-structured tagging

In this approach, which is potentially the most powerful for multimedia tagging, multiple users tag resources and interact within a community regarding the metadata and that metadata is well-formed and structured, typically based on a standardised metadata scheme. These are typically online systems with community capabilities (Yu et al., 2012; Parra-Arnau et al., 2013; Benna et al., 2012; Pan et al., 2012). For example, Video-Wiki (Blankinship and Mikhak, 2007) is an integrated suite of web applications for collaborative markup (with descriptive metadata, including real-time data streams) and remixing (by manipulating metadata) of video content. Facebook is a Web-based social networking service which, among other features, allows users to upload and tag videos, images and other posts, and share them with other users. It allows users to collaborate in the simple tagging of these resources (primarily just users featured or associated with the content and associated locations), comment on them and also provides live chat facilities. However, in many cases, users cannot access the multimedia resource without being ‘friends’ with the owner of the resource and therefore the collaboration is mostly limited to the owner’s connections in the social network (their community). As considered in the previous section, folksonomies are typically a collaborative-unstructured approach. However, some research has attempted to incorporate a formal ontology with collaborative folksonomy tagging in order to make it more structured. One way to do this is to derive it from the tags deployed within the system through data mining (Golder and Huberman, 2006b). Ulges et al. (2008) automatically tag videos by detecting high-level semantic concepts, such as objects or actions, in online portals like YouTube. Another method is to employ ontology seeding, which embeds an ontology into the system before the users commence tagging and typically asks the users for additional semantic information to ensure that the tags they contribute follow the conventions of the ontology (Golder and Huberman, 2006b). FolksAnnotation (Al-Khalifa and Davis, 2006), a system that extracts tags from del.icio.us and maps them to various ontology concepts, has helped to demonstrate that semantics can be derived from tags.

1.3 Challenges to collaborative multimedia content modelling

The amount and variety of multimedia content available on the Web has grown at an exponential rate in recent years, and thus the need to annotate multimedia content has become ever more imperative in order for users to be able to access these multimedia resources effectively (Money and Agius, 2008). Recent research in multimedia annotation approaches has witnessed an increasing interest in collaboration, not least due to the ability of Web 2.0 to support communities through the reuse and amalgamation of different Web services to provide rich application experiences (Lau and Chien-Sing, 2012; Das et al., 2012). For example, YouTube integrates video streaming and forum technologies with AJAX to support video-based communities. In this section, the research literature was reviewed to examine challenges to collaborative multimedia content modelling. The challenges were organised under criteria from (Jain et al., 1999) which are traditionally used for data clustering but are adapted here for use in the context of content modelling. Clustering is defined as the unsupervised classification of patterns into groups based on similarity (Jain et al., 1999). In this sense it is very similar to content modelling where concepts are tagged, for example, as objects or events based on their semantic similarity. These criteria are adopted in the content modelling context as:

- *Model stability*: In data clustering, stability is the degree by which the learning rate tends toward zero after a finite number of learning iterations and no pattern changes their clustering thereafter. For content modelling, this definition was adapted to mean the degree by which the model stabilises after a finite number of metadata additions so that no further model elements are required in order to achieve full closure of the semantic gap. Therefore, this is a measure of the degree of completeness of relationships between the semantics of the content and the metadata of the content model.
- *Model plasticity*: In data clustering, this is the ability of the algorithm to adapt to new data without having to re-cluster. For content modelling, this definition was adapted to mean the degree by which additional tags may be incorporated within the model. Therefore, this is a measure of the degree of extensibility.
- *Model accuracy*: In data clustering, this refers to the degree with which clustering classes reference ground truth classes. For content modelling, this definition was adapted to refer to the degree of correlation between the real world and the associated metadata; that is, the goal of the model. Therefore, this is a measure of the degree of correlation.

Table 1.1 presents a summary of all challenges along with a brief definition of each challenge. Figure 1.2 presents how the challenges are organised within the above categories.

Table 1.1: Challenge definitions

Category	Challenge	Definition
Stability	Pattern stability	To ensure a stable pattern for the metadata that conforms to the pattern of the tags created by the users.
	Identity awareness	To understand which users have authored or updated specific elements of the metadata, to know which users are presently active, or to be able to contact users.
Plasticity	Synonym control	To connect tags to their synonyms.
	Connectedness	To implement the full extent of relationships between tags in a way that makes sense to human users.
	Metadata propagation	To be able to tag a group of media streams at the same time/ apply a tag to a group of media streams.
Accuracy	Tag expressiveness	To spell check tags and connect them to their synonyms and also clarify the type of tag they are.
	Tag-based ranking	To provide means to rank the search results based on the tags that are used to annotate the media and the strength of these tags in each media.
	Inter-referential awareness	To provide means for the users to refer to specific elements of the content or the associated metadata during collaborative communication with others such that this reference may be identified and understood by all parties.
	Semantic awareness	To provide facilities for users within the community to be informed about the kind of changes that have been made to the metadata between revisions and which elements of the metadata have been authored by which users.

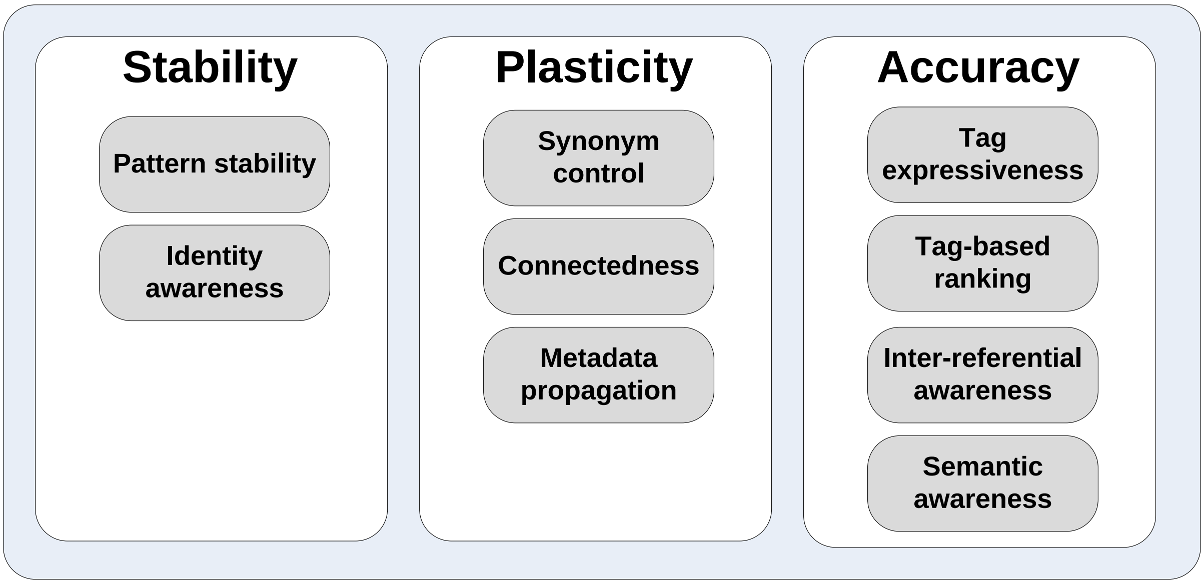


Figure 1.2: The categorisation of challenges of multimedia content modelling communities

1.3.1 Model Stability

The challenges explained here deal with how stable the resulting model is in regards to changes. Two challenges fall under this category: pattern stability and identity awareness.

The identification of patterns found within the tags that are created by users can help to enhance the retrieval process. For example, for a photo, users tend to specify the date taken, location and the visible people or landmarks. Recognising such *pattern stability* and encouraging users to conform to these patterns through, for example, suggested or mandatory fields can help to better direct the tagging process, subsequently producing well-formed tags (Hsu et al., 2012). Pattern stability ensures that the metadata conforms to the pattern of the tags created by the users. While there may be some loss of creativity and comprehensiveness in the annotation process, the usability of the tags within the retrieval process is improved. Consequently, defined structures can be enforced to ensure pattern stability across the community such that tag creation is controlled. Research has shown that although users demonstrate great variety regarding what keywords and phrases they use within tags and how frequently they use them, stable, aggregate patterns are emerging that can be used to help structure the tagging process (Golder and Huberman, 2006a; Hongbo et al., 2013; Ahmed and Guha, 2012). Ontologies also help provide pattern stability and one way to use a formal ontology with collaborative tagging is to derive it from the tags deployed within the system through data mining (Golder and Huberman, 2006b; Strohmaier et al., 2012). Ulges et al. (2008) present a system that automatically tags videos by detecting high-level semantic concepts, such as objects or actions, through using videos from online portals like YouTube as a source of training data, while tags provided by users during upload serve as ground truth annotations. Another method is to employ ontology seeding, which embeds an ontology into the system before the users commence tagging and typically asks the users for additional semantic information to ensure that the tags they contribute follow the conventions of the ontology (Golder and Huberman, 2006b). FolksAnnotation (Al-Khalifa and Davis, 2006), a system that extracts tags from del.ici.ous and maps them to various ontological concepts, has helped to demonstrate that semantics can be derived from tags. Queveo.tv (Barragáns-Martínez et al., 2010) is a recommendation system that uses tags added by users for various videos to build user and item tag clouds that allow for creating a folksonomy of the tags used by users to arrive at a semantic pattern for recommendations. However, before any ontological mapping can occur, the vocabulary must usually be converted to a consistent format for string comparison.

Studies (e.g. Razikin et al., 2011; Golder and Huberman, 2006b; Dihua et al., 2012; Allam et al., 2012) have shown that different users adopt different perspectives on how multimedia content should be modelled. Consequently, in order to understand which users have authored or updated specific elements of the metadata, to know which users are presently active, or to be able to contact users, *identity awareness* is also required. This includes self-awareness such that users may identify their own changes and activity. Identity awareness aims at identifying which users have authored or updated specific metadata and which users are presently active. Patterns and profiles can be used here to both raise awareness about users and increase collaboration standards between users (Gombotz et al., 2006; Carminati et al., 2012). Several researchers (Kim et al., 2011; Lipczak et al., 2012; Mezghani et al., 2012; Schoefegger and Granitzer, 2012; Schöfegger et al., 2012) have built collaborative user models (profiles) by leveraging user-generated tags. They take into account positive and negative tags the user has entered and from these deduce the user's likes and dislikes. Taagle (Maniu and Cautis, 2012) is a social tagging system for a social network whose tags reflect friendship, similarity, tagging behaviour or trust. Thornton and McDonald (2012) argue for the use of category systems in collaborative tagging which have traditionally been created by small groups with authority over system design. Moreover, there have been several community configuration management tools developed, such as Palantir (Sarma et al., 2003), a workspace awareness tool that informs a user of which other users have changed artefacts, calculates the severity of the changes, and graphically displays the information in a configurable and generally non-obtrusive manner. As with logging and revision control systems, however, community configuration management tools have yet to be adapted to and implemented in multimedia communities, where additional, distinct activities must be catered for, such as updating different content features and media streaming.

1.3.2 Model Plasticity

The challenges in this category focus on the extensibility of the model created through collaborative content modelling. This category includes the following challenges: synonym control, connectedness and metadata propagation.

One of the challenges in collaborative content modelling is that tags are highly individualistic and different users often use a variety of terms to describe the same concept (Lee and Yong, 2007; Dellschaft and Staab, 2012). This makes searching problematic since users may only search for one synonym but require results for the entire group of synonyms (Fuentes-Lorenzo et al., 2013). In contrast to traditional subject indexing, metadata in communities is generated not only by experts but also by creators and consumers of the content. Usually, freely chosen keywords are used instead of a controlled vocabulary (Voss, 2007). Batch et al. (2011) pointed out that using free words chosen by users leads to tags being inefficient in terms of meanings and semantic interlinking capabilities and therefore present MTag (Batch et al., 2011) as a model for collaborative medical tagging. Consequently, such a lack of *synonym control* makes searching problematic since users may only search for one synonym but require results for the entire group of synonyms. Synonym control connects tags to their synonyms.

Exploiting the full *connectedness* of the tags is a further challenge. Connectedness implements the full extent of relationships between tags in a way that makes sense to human users. The full extent of relationships between tags that make sense to humans is difficult to develop within tag-based systems and thus they are often not implemented (Uddin et al., 2013; Zhang et al., 2013; Zhang et al., 2012). For example, if a user searches for 'swimming', the results should ideally include all terms related to swimming, not just synonyms but also related concepts such as medals, Olympic events and so on. Language tools to identify synonyms, acronyms and relationships have sought to address the above issues. One of the most popular tools used is WordNet (Fellbaum, 1998), which is an English semantic lexicon. It groups English words into sets of synonyms, provides general definitions and acronyms, and records the various semantic relations between terms. At the core of WordNet is the use of synsets (also known as synonym rings). A synset is a group of words that denote the same concept and are interchangeable in many contexts. Each synset is linked to other synsets by means of conceptual relations. Some approaches build on the notion of a text-based synset for image annotation, such that a visual synset is used as an organisation of images which are visually-similar and semantically-related. Each visual synset then represents a single prototypical visual concept with an associated set of weighted annotations (Zheng et al., 2008; Tsai et al., 2011). WordNet has been used to solve tagging issues and develop retrieval systems that utilise folksonomies, e.g. TagPlus (Lee and Yong, 2007) retrieves from Flickr by using WordNet to correct, identify and group tags generated by users to improve the relevance of the presented results. Zhu et al. (2012) use WordNet on datasets from Del.icio.us for social tag recommendation. San Pedro et al. (2011) propose an approach that uses content duplications in various video sharing Websites to reach connectedness by discovering the tags used by a variety of users to annotate the same content. However, such systems are add-ons to existing systems and not integrated into the community, which limits the breadth and depth of their functionality.

Another challenge is that of *metadata propagation*, i.e. being able to tag a group of media streams at the same time or apply a tag to a group of media streams. Often it is the case that multiple, related media streams require tagging, which may be carried out simultaneously or sequentially. Such streams share many common content features and thus tagging becomes a repetitive process involving a lot of redundancy. This problem is exacerbated in collaborative modelling systems where there are multiple users carrying out the tagging. For example, photos of a certain occasion will often be taken by many users who were present and these photos will contain common content features, such as the people present and the location. If the metadata could be propagated or inherited by all related photos once one, or a small number of, users had tagged them, user effort would be greatly reduced while also greatly improving the consistency of the tags (William, 2006; Choi and Suh, 2013; Seo and Lee, 2013; Yang et al., 2012). Ning et al. (2011) present a hybrid probabilistic model which integrates low-level image features and high-level user provided tags to automatically tag images. For instance, with two groups of images, one tagged with "sky" and "tree" and the other tagged with "tree" and "grass," a new group of images tagged with only "grass" could be predicted to have the tag "sky" even though "grass" and "sky" have never been tagged in the same image by any users.

1.3.3 Model Accuracy

The challenges of this section deal with the accuracy of the model created through collaborative content modelling. In other words, they focus on how accurately the model relates to the real world semantics. The challenges of this category are: tag expressiveness, tag-based ranking, inter-referential awareness and semantic awareness.

Tagging is so widely used now that users have become comfortable with it to the point where they do not give it much conscious thought and therefore do not consider the clarity, structure and subsequent usability of the tags they create (Lee and Yong, 2007; Tourné and Godoy, 2012; Bonchi et al., 2012; Figueiredo et al., 2013). Consequently, *tag expressiveness* is a non-trivial challenge aimed at spell-checking tags and connecting them to their synonyms and also clarifying their type. For example, unstructured text in tags often results in having many tags that are difficult to interpret and relate to other tags by anyone other than the tag author (Golder and Huberman, 2006a; Djuana et al., 2012; Chatzilari et al., 2012). Similarly, incorrectly spelled keywords within the tags can cause retrieval problems and lead to orphaned content that can only be retrieved when the query also happens to contain the same misspellings. Acronyms also prove problematic and can reduce the usability of the tags when they are not commonly accepted and are decided by the content author alone. Consequently, users may search for different acronyms or search for the full phrase instead, both of which would fail to retrieve the required content. Also, users have different intentions while tagging, hence not all the tags available are related to the content of the annotated item. This can include tags such as self-references and personal tasks ("my husband", "to do list") or may be expressing subjective opinions and qualities ("nice book", "dark movie") (Cantador et al., 2011).

Tag-based ranking aims at ranking search results based on the tags that are used to annotate the media. Studies show that tag-based rankings produce more relevant results than traditional rankings and clusterings (Firan et al., 2007; Bellogín et al., 2013; Zhang et al., 2013; Jian and Qun, 2012; Doerfel et al., 2012), as demonstrated by the recommendations of online music sites and communities such as Yahoo Music and Last.fm. Zhang et al. (2009) propose a video blog management model which comprises of automatic video blog annotation and user-oriented video blog search. For video blog annotation, they extract informative keywords from both the target video blog itself and relevant external resources. As well as this semantic annotation, they perform sentiment analysis on comments to obtain an overall evaluation. For video blog search, they present saliency-based matching to simulate human perception of similarity and organise the results by personalised ranking and category-based clustering. The HeyStaks (McNally et al., 2011) system is a social Web search system designed to help users collaborate during Web search tasks and it combines collaborative recommendation techniques with main stream search engines. It relies on users' click-throughs, voting, sharing and tagging/commenting to rank the results from a Web search. Tagging is considered a more reliable indicator of interest than a simple result click-through or vote.

Inter-referential awareness considers how a user may refer to specific content elements or associated metadata during collaborative communication with other users (Chastine et al., 2006; Qin et al., 2012), both in asynchronous and synchronous environments. While verbal references may be sufficiently effective, though inefficient, for metadata references, (Fu et al., 2010) they are less suitable for references to content within the media stream (Vijay and Jacob, 2012) since users may be attracted to the wrong features during playback or they may become engrossed in an event such that they do not notice certain objects appearing (Pea et al., 2006). In the physical world, looks, nods and hand gestures serve to focus attention effectively and thus one would expect suitable surrogates to be available within a collaborative environment. Both Microsoft Office and Adobe Acrobat enable users to add basic annotations to their work so that users can draw the attention of other users. Similar tools catering for audio-visual files in collaborative multimedia communities would be invaluable.

When different users work collaboratively to model the same media content, different versions of the same content model may be generated. Consequently, it is important for users within the community to be informed about the kind of changes that have been made to the metadata between revisions and which elements of the metadata have been authored by which users (Papadopoulou et al., 2006; Isotani et al., 2013) in order to correct the metadata, continue their own content modelling, or rethink their own decisions and modelling approach if necessary (Shen and Sun, 2005). In collaborative content modelling systems, the changes made to a content model are not limited to just changes in phrases and keywords but more importantly relate to full semantic content features such as objects, events and the relationships between them. *Semantic awareness* refers to user facilities regarding revisions to the metadata and user authorship. To help provide semantic awareness, many logging and revision control approaches have been proposed, such as

IceCube (Kermarrec et al., 2001), which enables general-purpose log-based reconciliation, where logs of alterations are combined into a single merged log and, by observing object and application semantics, are ordered in such a way as to minimise conflicts. Another method proposed (Tang et al., 2011) has been group profiling, where all the actions of users are gathered, similar to social media profiles, and grouped together around a concept such as a search query. Some researchers have implemented privacy protection methods for social tagging (Parra-Arnau et al., 2013; Parra-Arnau et al., 2012). Such approaches have yet to be adapted to the semantics of multimedia content models and incorporated into multimedia communities where they must work with content, context and language. Existing multimedia communities, such as YouTube and Flickr, either do not allow editing or they do not control the changes and conflicts during multimedia annotation.

1.4 Research Aim and Objectives

The aim of this research is to explore collaborative modelling of multimedia content and in particular how effective existing tools are. Through the consideration of both the shortcomings of existing tools and the future challenges faced with collaborative modelling, design and develop a framework and implement an online service that caters for MPEG-7 content modelling communities. In order to achieve the research aim, the following research objectives are pursued:

- O1. Design a user experiment in order to study users' tagging behaviour with existing multimedia tagging tools and identify any relationships between such user behaviour. The results will be used to construct a conceptual model of user behavior.
- O2. Design and develop a framework for MPEG-7 content modelling communities (MC^2) based on the conceptual model of user behaviour developed in pursuit of O1. The resulting framework will be called MC^2.
- O3. Implement the online service based on the MC^2 framework developed in pursuit of O2 as a proof of concept. The MC^2 online service will be released for use by selected groups of users in preparation for pursuing O4.

O4. Validate the MC^2 framework developed in pursuit of O2 with the online service developed in pursuit of O3 through a repeat of the user experiment designed in pursuit of O1. The validated MC^2 online service will be released for public use.

Table 1.2 summarises the research objectives and their outcomes.

Table 1.2: Research objectives and outcomes

Objective	Outcome
O1	Conceptual model of user behaviour
O2	MC^2 framework
O3	MC^2 online service
O4	Validated MC^2

1.5 Research Methods

Table 1.3 shows the research methods which have been used to pursue each research objective.

Table 1.3: Research objectives and outcomes against research methods used

Objective	Outcome	Research Method
O1	Conceptual model of user behaviour	User Experiment Design, Grounded Theory
O2	MC^2 framework	UML Use Cases, Grounded Theory
O3	MC^2 online service	Rapid Application Development
O4	Validated MC^2	User Experiment, Grounded Theory

In pursuit of ***Objective 1***, an experiment was designed during which a group of users were tracked whilst completing a series of tasks using four existing multimedia metadata tools. Throughout the experiment the users remained in constant communication with each other and their behaviour during all collaboration was recorded against the relevant multimedia content. All data collected during the experiment, i.e. metadata created by the four tools, the collaboration transcripts, and their responses to post-experiment interviews, were analysed using grounded theory (Corbin and Strauss, 2008) and a fuzzy cognitive model of user behaviour is developed. A grounded theory is defined as theory which has been “systematically obtained through social research and is grounded in data” (Goulding, 1998). It represents and meets four central criteria: fit, understanding, generality and control (Corbin and Strauss, 2008): the theory fits the substantive data, the theory is understandable to all involved in the area of study, the theory is generally applicable in a variety of contexts, and the theory provides control with regard to action toward the phenomenon. Grounded theory methodology comprises of systematic techniques for the collection and analysis of data, exploring ideas and concepts that emerge through analytical writing (Charmaz, 2006). Grounded theorists develop concepts directly from data through its simultaneous collection and analysis (Matavire and Brown, 2008). In this way, it contrasts with the positivist scientific method and that of a priori hypothesis formulation in that the resultant theory is inductively derived from the data. Instead, grounded theory sets out to find what theory accounts for the research situation as it is. In this respect, it is like action research such that the aim is to understand the research situation; that is, to discover the theory implicit in the data. Consequently, grounded theory takes a case rather than a variable perspective, such that different cases are taken to be wholes, in which the variables interact as a unit to produce certain outcomes. A case-oriented perspective tends to assume that variables interact in complex ways, in contrast with simple additive models, such as ANOVA, with main effects only, thus a comparative orientation is supposed. Cases similar on many variables but with different outcomes are compared to see where the key causal differences may lie and cases that have the same outcome are examined to see which conditions they all have in common, thereby revealing necessary causes. Categorisations tend to be drawn from users themselves and thus the focus is on making implicit belief systems explicit. The process of using grounded theory starts with *open coding* which includes theoretical comparison and constant comparison of the data, up to the point where conceptual saturation is reached. This provides the *concepts*, otherwise known as *codes*, that will build the means to tag the data in order to properly *memo* it and thus provide meaningful data (dimensions, properties, relationships) to form a theory. Conceptual saturation is reached when no more codes can be assigned to the data and all the data can be categorised under one of the codes already available, with no room for more codes. In order to assist with the analysis and induction of the grounded theory, a visualisation stage has been included to follow memoing, the outcome of which is the fuzzy cognitive model of conceptual user behaviour.

In pursuit of ***objective 2***, UML use-cases were used with grounded theory to produce a grounded design of the MC^2 framework (Object Management Group, 2011) based on the Conceptual Model of User Behaviour produced by objective 1. Although grounded theory is not normally used for such purposes, Glaser claims the dictum "all is data" (Glaser, 1998 p. 9; Kelle, 2005) and, thus, according to his approach to grounded theory, it can be applied and adapted for *any* research. Indeed, grounded theory has been used successfully in unconventional cases previously, such as software engineering (Carver, 2007) and creating process models (Tian, 2006). Furthermore, the outcome of grounded theory is often one or more conceptual models. Since both a use case diagram and an architectural framework are both forms of a conceptual model in essence, it is reasonable to consider them both valid outcomes of a grounded theory process. Evident requirements were considered those which were not explicitly specified but were evidently necessary in order for the framework to be fully functional. For example, there is a requirement for users to upload media so that it is available within the framework.

In pursuit of ***objective 3***, an online service is implemented as a proof of concept of the MC^2 framework developed in pursuit of objective 2 using Rapid Application Development (RAD). It involves iterative software development and software prototyping. It is a merger of structured techniques with prototyping techniques to accelerate software development (Gerber et al., 2007; Maurer and Martel, 2002).

Finally, in pursuit of ***objective 4***, the MC^2 online service developed in pursuit of objective 3 is validated through the same user experiment designed in pursuit of objective 1 with the same group of users. All data collected during this new experiment are likewise analysed using grounded theory.

1.6 Thesis Outline

Chapter 2 presents the development of the conceptual model of user behaviour visualised as a fuzzy cognitive map. The model is developed through a series of user experiments with existing multimedia tagging tools whose aim is to unravel the relationships that exist between context and user tagging behaviour using grounded theory.

Chapter 3 presents the development of the MPEG-7 MC^2 framework and online service for collaborative content modelling. The MC^2 framework is developed from the fuzzy cognitive model of user behaviour using use cases and grounded theory. The MC^2 online service is implemented using rapid application prototyping both as proof of concept of the MC^2 framework and as a service to the experiment participants.

Chapter 4 presents a walkthrough of the functionality and a validation of the MC^2 online service. The MC^2 online service is validated, first, as a tool and, secondly, for its functionality with the same group of users and through the same user experiments that led to the development of the fuzzy cognitive model of user behaviour. The results obtained during the validation are compared against those of the multimedia tagging tools that were used in the initial experiments.

Chapter 5 summarises the research reported in the thesis and the contributions made by this research, namely the fuzzy cognitive model of user behaviour and the MC^2 framework and online service. The chapter presents emergent research and development threads with the research contributions, ranging from the immediate to the short-term to the long-term.

Chapter 2: A Conceptual Model of User Behaviour

While many (typically isolated) approaches have been proposed to support collaborative metadata creation, very little empirical research exists which seeks to understand how users actually tag multimedia. Gaining such understanding would enable progression towards a system that would not sacrifice metadata for the sake of collaboration (or vice versa) and thus would provide improved community support along with comprehensive metadata. The aim of this chapter is to develop a conceptual model of such user behaviour and visualise it as a fuzzy cognitive map. In order to achieve this aim, an experiment was designed during which usage data was collected and then analysed using grounded theory and the results were visualised as a fuzzy cognitive map of user behaviour.

2.1 Experiment Design

An experiment was undertaken with 51 users where these users were asked to undertake a series of tasks using four existing multimedia metadata tools and their interactions were tracked. The users were chosen from a diverse population in order to produce results from typical users. The users were aged between 20-46 at the time of the experiment, with a distribution of 45% female and 55% male. The users were unsupervised, but were communicated with other users via an instant messaging application, e.g. Windows Live Messenger, so that transcripts of all conversations could be recorded for later analysis (sample attached as Appendix VII). These transcripts contain important information about the behaviour of users in a collaborative community and are considered to contain metadata information if they contain comments explicitly related to the content. To design the experiment, the fundamental multimedia content types and content features that should be tagged were taken into account and the collaborative user groups within the experiment were organised accordingly. The users' ethical approval was attained through a consent form attached as Appendix II. Next the experiment design is presented and the research method used for analysis is discussed. Results are presented in Section 2.2.

2.1.1 Materials

The materials used in the experiment are divided into three categories of; multimedia content types, content features and experiment tools.

2.1.1.1 Multimedia Content Types

In order to ensure that the experiment covers a broad range of content, it is necessary to acknowledge the various multimedia content types available. There are many different categorisation criteria that may be used. For example, the multimedia resources may be categorised based on content format, such as audio, video and image (Qiu et al., 2008), or based on Wikipedia topic categories (Mishra et al., 2010). However, neither relate directly to the actual content of the multimedia resources, while the latter also has the disadvantage of yielding an unmanageable number of categories. Thus, multimedia resources are categorised according to their semantic content into the following four areas (Smoliar and HongJiang, 1994; Kumar and Tomkins, 2010; Otsuka et al., 2005):

- **Personal:** This type of content is personal to users, e.g. videos of family, friends and work colleagues. Content is typically based around the people, occasion or location. This type of content is mainly found on personal storage devices and typically uploaded onto sharing repositories including YouTube, Facebook and Flickr.
- **Business:** This type of content is created and used for commercial purposes. It mainly includes videos created for advertising and promotion, such as video virals. This type of content is mainly held in commercial repositories, e.g. WISTIA[1] and StartupTV[2].
- **Academic:** This type of content serves academic purposes, e.g. teaching and learning or research. This type of content can be found on university websites and professional

[1] http://www.wistia.com
[2] http://www.startuptv.co.uk

academic websites such as TED[1]. This content is also distributed through publishing software like iTunes.

- **Recreational:** This type of content is created and used for purposes other than the above and includes faith, hobby, amusement or free-time-based content. Recreational content can be found on entertainment websites, e.g. record labels, production studios, some personal websites, church websites, and websites such as YouTube and Vimeo[2].

The above categorisation can be seen to represent the majority of multimedia content, ensuring that users within the experiment are tagging a representative sample of relevant multimedia content.

2.1.1.2 Content Features

Tagging of multimedia resources involves the recognition of content features exhibited by the multimedia resources and the description of these as metadata. In order to ensure that the experiment covers a broad range of content features, it is necessary to acknowledge the various multimedia content features that may exist. These features can be low-level AV features (Li et al., 2011), e.g. colour structure, colour layout, and scalable colour representing colour features, along with the homogeneous texture and edge histogram representing texture features (Hyun-seok et al., 2010), or they can be high-level AV features consisting of key frame descriptions (Ding et al., 2010), or high-level semantic features of the multimedia resource. These features are highly utilised in querying and retrieval (Zhou et al., 2010; Cui et al., 2010; You et al., 2010) and can be categorised into five areas (Agius and Angelides, 2007; Agius and Angelides, 2009):

- *Objects:* People, animals, and inanimate objects.
- *Events:* Visual or aural occurrences within the video, e.g. a car chase, a fight, an explosion, a gunshot, a type of music. Aural occurrences include music, noises and conversations.

[1] http://www.ted.com
[2] http://www.vimeo.com

- *Spatiotemporal locale:* Positions the above in content or media time and space. Thus, the spatiotemporal locale may be a semantic location or time, e.g. London, morning, Middle Ages, or a precise clock date-time or geographic positioning such as longitude and latitude, as well as a content time-point (start-time, end-time, duration).
- *Properties:* Properties of the above.
- *Relationships:* Relationship between the above which may be temporal, spatial or semantic. Typical semantic relations include *causer* (causes another event or object to occur), *user* (uses another object or event), *part* (is part of another object or event) and *specializes* (a sub-classification of an object or event).

Together, these five types of features can be considered to create a comprehensive description of a multimedia resource. However, not all features are available in all types of content and some content will be richer in terms of some features over others. For example, personal content is typically rich in people, location and certain types of events (such as celebrations), but aural events such as noise, and semantic relations such as part, are less commonly exhibited features.

2.1.1.3 Experiment tools

The multimedia tagging tools used in the experiment are:

- **YouTube for Videos:** This tool provides a community for sharing video content on the Web. YouTube enables users to upload their videos, set age ratings for the videos, enter a description of the video, and also enter keywords.

- **Flickr for Images:** This tool provides a means for photo enthusiasts at all levels to upload and share their photos in the original quality. It also provides tools for tagging the pictures. Flickr also enables users to perform simple manipulations on their photos such as rotating and cropping.

- **del.icio.us for websites that contain AV (audio-visual) content:** This tool enables Internet users to bookmark and tag pages they come across while browsing through the Internet and share these with other users. They can also add comments and descriptions to their tagged pages. AV content is now commonly found on various websites, e.g. blogs, and therefore del.icio.us provides a means to tag these pages and add descriptions of the AV content found on the pages.

- **COSMOSIS for AV content:** This system provides the means to model the content based on the COSMOS-7 MPEG-7-based metadata scheme. With this system, users can model videos and define the semantics of its content such as objects, events, temporal relations and spatial relations.

The participants also used instant messengers for the live communication during the experiment. The choice of live messenger was left to the users. The users utilised three different instant messengers; Windows Live Messenger, Yahoo Messenger and Skype.

2.1.2 Participants

The 51 users of the experiment were aged between 20 and 46 at the time of the experiment, with a distribution of 45% female and 55% male. The users were chosen from a diverse population with various levels of competency with tagging and multimedia tagging. As the aim of the experiment is to observe and discover tagging behaviour of average user, it was necessary that the participant had various levels of competency and not be proficient in tagging and multimedia content modelling.

The users' backgrounds were diverse; the participants included users with academic backgrounds that consisted of academics, researchers and PhD students. There were also users with backgrounds in marketing, advertising and sales. There were also users that had backgrounds in arts such as photographers, graphics designers, musicians. There were also certain users that had a background in online collaborative environments such as digital gamers and bloggers. All the users were recruited through personal contacts who would have been willing to participate for an hour in the experiment and had access to the internet.

2.1.3 Procedure

Users were given a series of tasks (see Appendix I), requiring them to tag the AV content using a selection of unstructured (folksonomy) and structured (MPEG-7) multimedia tagging tools that together represented the tagging approaches discussed earlier and presented in Figure 1.1. All tools were augmented with external communication facilities outside of the tool (discussed below) to ensure that users were able to fully collaborate when using the tools, and were not limited to the pre-existing (if any) collaboration support provided by the tool. The four tools used were: YouTube, Flickr, del.icio.us and COSMOSIS. The first three tools may be considered popular folksonomy tools and were chosen to cover videos, images and audio/video content respectively. Strictly speaking, YouTube is missing the aggregate front end that Flickr and del.icio.us have which formally make them folksonomy tools, but YouTube does contain all other folksonomic elements and thus is the closest available folksonomy tool for video. It is also an extremely commonly used tool and therefore familiar to a wide range of users. COSMOSIS is an advanced, structured tagging tool which supports the MPEG-7 standard.

Users were assigned to groups (12-13 per group), one for each of the four different multimedia content types identified earlier, but were not informed of this. Users were assigned to one of four groups based on their background and familiarity with the material. Users with an academic background (academics, researchers and PhD students) were assigned to the academic group, while users with either a marketing, advertising or sales background were assigned to the business group. Users in the recreational group came from backgrounds that included photography, graphic design, music, digital gaming and blogging. The remaining users were assigned to a personal group as this content type is suitable for all users.

Within these category groups, users worked together in smaller experiment groups of 3-6 users to ease the logistics of all users in the group collaborating together at the same time. Thus, let U be the set of all users taking part in the experiment and $|U| = 51$. Then, G_i is a set of users assigned to category group i, where $1 \leq i \leq 4$, $12 \leq |G_i| \leq 13$, $G_i \subsetneq U$ and $\bigcap_{i=1}^{4} G_i = \emptyset$, such that G_i partitions U; and H_j^i is a set of users assigned to experiment group j within category group i, where $3 \leq |H_j^i| \leq 6$, $H_j^i \subsetneq G_i$ and $\bigcap_{j=1}^{|H_j^i|} H_j^i = \emptyset$, such that H_j^i partitions G_i. Members of the same group were instructed to communicate with other group members while they were undertaking the tasks using an instant messaging application, e.g. Windows Live Messenger. Thus, group membership took into account user's common interests and backgrounds since this was likely to increase the richness and frequency of the communication. The importance of user communication during the experiment was stressed to users.

Given the common content feature differences between the multimedia content types (personal, business, academic, recreational), the videos for each category group were noted to differ in which features they were rich in, with other features also exhibited. The personal category group (Group 1) was asked to use their own videos, the business category group (Group 2) was provided with business-oriented videos, the academic category group (Group 3) was provided with videos of an academic nature, and the recreational category group (Group 4) was provided with a set of recreational videos. Each user was required to tag the content of 15 images in Flickr, 10 web pages containing AV content in del.icio.us, and 3-5 minutes worth of videos in YouTube and COSMOSIS (either one 5 minute long video or a number of videos that total 5 minutes together). This ensured that users need not take more than about an hour to complete the tasks, since more time than this would greatly discourage them from participating, either initially or in completing all tasks. At the same time, the video duration, number of images and web pages are sufficient to accommodate meaningful semantics. Users did not have to complete all the tasks in one session and were given a two week period to do so. YouTube, Flickr, del.icio.us and COSMOSIS metadata and collaborative communication transcripts were collected post experiment.

After the users had undertaken the required tasks, a short, semi-structured interview was performed with each user. The focus of the interviews was on the users' experiences with, and opinions regarding, the tools. The questions were carefully selected based on the challenges so that some concerned tagging while others addressed context. Typical questions included which tool the users found easiest and most functional to use, which tags were used most when tagging and describing the AV content, which aspects the users felt important when tagging, and if they felt certain aspects were more important for different types of content, and additional features the users would have liked to have seen in the tools (see Appendix V for interview questions and Appendix VI for sample answers).

2.1.4 Research Ethics

Ethical approval is required for research projects that involve human participation. Before approaching the participants, the Brunel University's Code of Research Ethics and General Ethical Guidelines and Procedures were consulted and an approval was obtained from the Research Ethics Review Board. The users' ethical approval was attained through a consent form attached as Appendix II.

2.1.5 Data Analysis: Grounded Theory

All of the data collected during the experiment, that is, the metadata created by the four tools, the collaborative communication transcripts, and the responses to the interview questions, were analysed using grounded theory. Figure 2.1 illustrates the steps which will be taken in this data analysis approach.

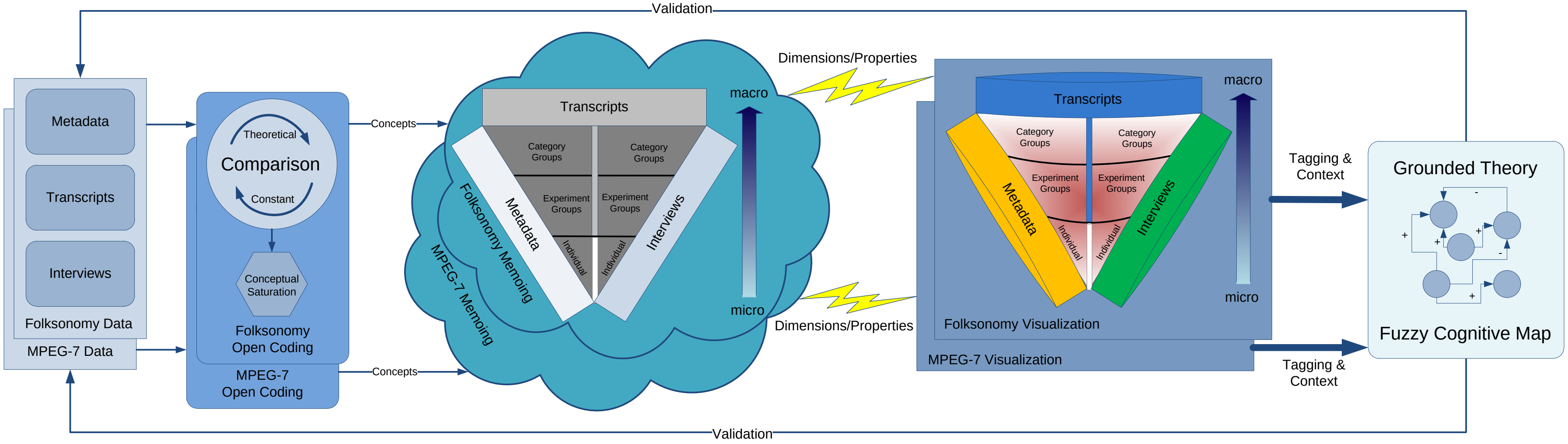


Figure 2.1: Grounded theory as applied to the collected data in this experiment

As can be seen in the figure, the metadata are gathered from the MPEG-7 and folksonomy tools, along with the collaborative communication transcripts and interviews, which form the basis of the open coding process. During the open coding process, this data goes through a cycle of theoretical and constant comparison until conceptual saturation is reached. The concepts generated by the open coding process are then used as a starting point for the memoing process. Thus, given the set of all tagging data, X, the set of all interview data, I, and the set of all communications transcripts, T, all the data derived from the experiment can be represented by $\mathbb{D} = X \cup I \cup T$. Open coding may therefore be considered to derive a set of codes associated with the data $\{C_1, C_{2,} \dots C_n\}$. Each C is a pair (c, D) where c is a code and $D \subsetneq \mathbb{D}$. All data is constantly compared such that conceptual saturation is reached once, or before, the full domain of comparison is considered; that is, $X^2 - \Delta_D \cup I^2 - \Delta_I \cup T^2 - \Delta_T \cup X \times I \cup X \times T \cup I \times T \cup X \times I \times T$, where Δ is the diagonal, and thus each data item is compared to every other data item but not itself. For example, some c reflected the key content features of AV content, which the experiment considered to be: Objects (people, animals, and inanimate objects), Events (visual or aural occurrences within the video, with aural occurrences including music, noises and conversations), Spatiotemporal locale (which positions objects and events in content or media time and space, i.e. semantic location/time, precise clock date-time, geographic position, or content time-point), Properties of the previous features, and Relationships between the previous features (temporal, spatial or semantic).

The memoing process is undertaken on three levels of analysis: individual, experiment groups, and category groups. Memoing is the theorising of ideas about fundamental codes and their theoretically coded relationships as they emerge during coding, collecting and analysing data, and during memoing (Glaser, 1998). This is represented as a pyramid in Figure 2.1 to show the increasing data for consideration that is adopted in this approach; that is to say, each level includes its own data as well as that of the levels below it. Each side of the pyramid shows the data source that is used in the memoing process: interviews, metadata and transcripts. As shown in the figure, transcript sources are not considered at the individual level since this data is not generated from the experiments (users always communicated within groups).

The memoing process commences on the individual level whereby the interview data from each user that takes part in the experiment is memoed independently. The next level of analysis considers the data of all the individuals who are in the same experiment group along with the interview data from the individual level of analysis. Following this, the interview data from each user category group (personal, academic, business and recreational) is considered in turn (along with that of the previous levels) so that the interview data from all the users who are assigned to the same category is memoed together to allow further groupings to emerge. Finally, all the interview data that is generated is considered together, as a whole. A similar process is then applied to the metadata generated by the user during the experiment. Then, the transcripts are processed. These are considered at the experiment and category group levels of analysis, as well as being memoed in aggregate. It is not possible to analyse the transcripts at individual level as the conversation took place in a group (the experiment group) so there is no individual record for each participant. Finally, all the interview, metadata and transcript data are considered in their entirety to derive any further dimensions and properties. At the conclusion of the memoing process, a set of memos associated with the codes and data $\{M_1, M_{2,} \dots M_n\}$ were derived. Each M is a triple (m, Γ, D) where m is a memo, $\Gamma \subsetneq C$ and $D \subsetneq \mathbb{D}$. The raw data from the metadata and the interviews were imported into NVivo[15], qualitative data analysis software, as documents so that it could assist with processes of open coding and memoing.

The visualisation stage follows the same approach as the memoing stage, whereby the data generated from each of the sources is processed through each of the three levels. The visualisation stage involves the visual representation of memos and their relations. The output is a set of content and usage aspects which together encapsulate the content features that emerge from the

[15] http://www.qsrinternational.com/products_nvivo.aspx

experiments, how they were used, and how users would like to use them. These content and usage aspects enable the derivation of a grounded theory in the form of a fuzzy cognitive map. To ensure the appropriateness of the fuzzy cognitive map, it is validated by comparison against the original data and all of the above processes are repeated again (and again) as necessary to ensure it is a fair representation of the data.

The next section, 2.2, presents the outcomes of the visualisation and memoing processes.

2.2 Experiment Results

In order to progress collaborative multimedia tagging tools such that the underlying metadata scheme and collaboration support may work in synergy, it is necessary to understand how users behave when tagging and the context to their tagging. The knowledge of the frequency with which users actually employ tags, which tags were most and least used, corroborated by the collaborative communication transcripts and the answers users gave when interviewed, is essential for the design of effective and efficient collaborative multimedia tagging tools that seek to prioritise tags within the metadata core that are used more often by users and improve the collaborative context to effectively support tagging. To this end, this section presents the results of this experiment: first, the results relating to tagging, then the results relating to context are presented.

2.2.1 Tagging

Figure 2.2 and Figure 2.3 present the overall tag usage in the folksonomy tools and the MPEG-7 tool, respectively, for each group, while Figure 2.4 depicts the overall tag usage for folksonomy and MPEG-7 tools for all groups combined. All three figures are derived from the application of the grounded theory method, taking into account both the empirically collected metadata and the responses to the interview questions. The analysis of the experiment data (memoing stage) considered the data within the sub-categories of the five content feature categories presented earlier so as to provide greater detail. Consequently, rather than memoing just objects, they were memoed as animals, people and inanimate objects, for example. Objects and events were easily identified in the metadata; however, the same was not true of the other three content feature categories. Spatiotemporal locale was split into time and location. In the folksonomy tools, most of the time tags were semantic times, such as morning or summer, with few specific dates. In COSMOSIS, however, time was mainly tagged as time-points (start-time and duration) for various entities. Locations in folksonomy tools were easily spotted as a tag was memoed as location when it was clearly a location such as a seaside. In COSMOSIS, the locations were tagged using SemanticRelation – locationOf; therefore, each location tag was explicitly defined in the metadata using a semantic relation. The tags memoed as properties in the folksonomy tools were the adverbs, adjectives and other descriptors used to further describe an object, e.g. for *bright yellow sun*, *bright* and *yellow* were considered properties. In COSMOSIS, the properties were defined by an instance of SemanticRelation - propertyOf. For both locationOf and propertyOf, the semantic relation tag does not have any descriptive benefits per se; it is merely one way to associate property and location metadata with other elements. The user could, in fact, add location and property metadata without using semantic relations, as is the case with folksonomy tools. Therefore, the benefits of the locationOf and propertyOf semantic relations can be seen to derive exclusively from structuring the metadata rather than the representation of semantic information. Consequently, for comparability, the semantic relation used for these tags have not been included in the total count of the semantic relations. Furthermore, while in MPEG-7 (COSMOSIS) all relations are explicitly defined and structured, folksonomy tools do not provide such a facility. The relations are always implicit in the metadata as the tools do not enable the tags to be input in a structured manner. Hence, there is great variety and ambiguity regarding what relations are actually represented, making it not only difficult to identify a specific type of spatial, temporal or semantic relation, but also to identify comparable relations between folksonomy and MPEG-7 tools. As a result, only instances of spatial, temporal or semantic relations were identified when memoing, rather than the specific spatial, temporal or semantic relations that were used. As can be seen from the figures, tag usage results are consistent across the groups. Comparing the group results with the overall results in Figure 2.4, the pattern of tag usage remains the same which suggests that the overall results for the folksonomic tools and the MPEG-7 tool are suitably representative of the results in the categories.

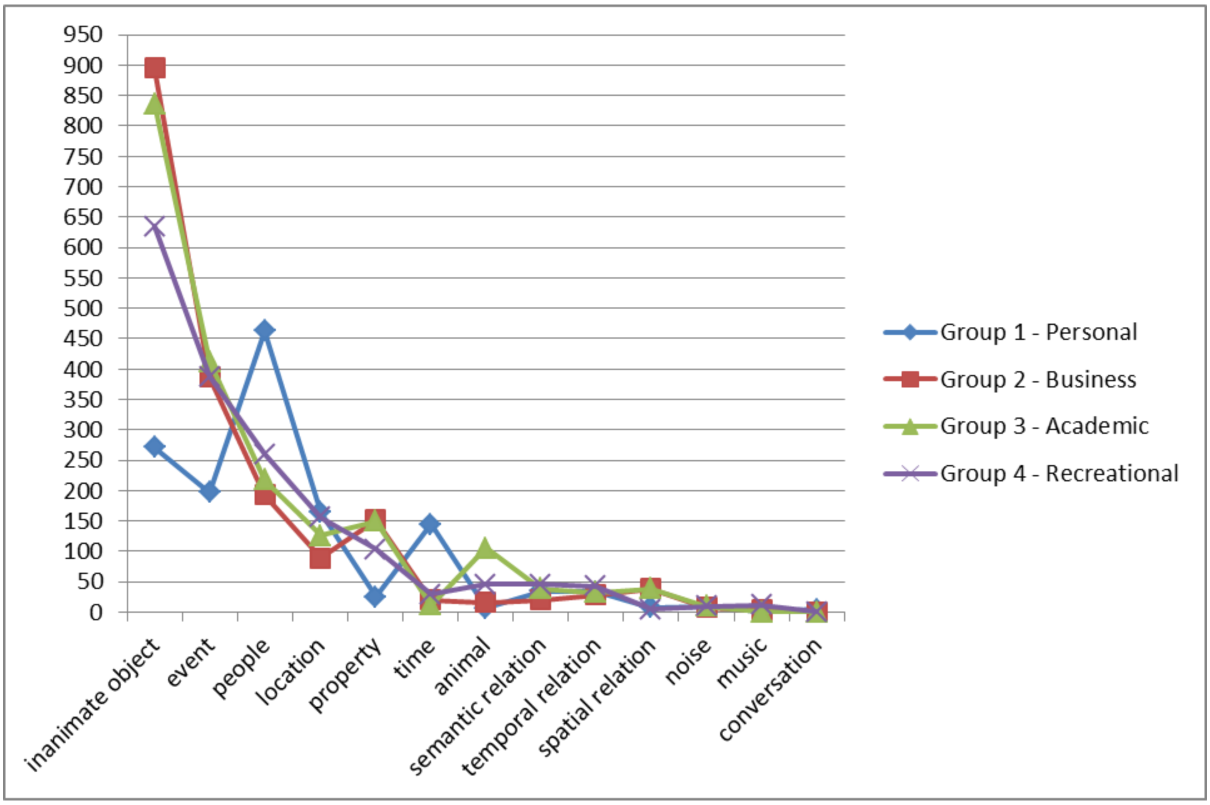


Figure 2.2: Overall usage of tags in folksonomy tools per group

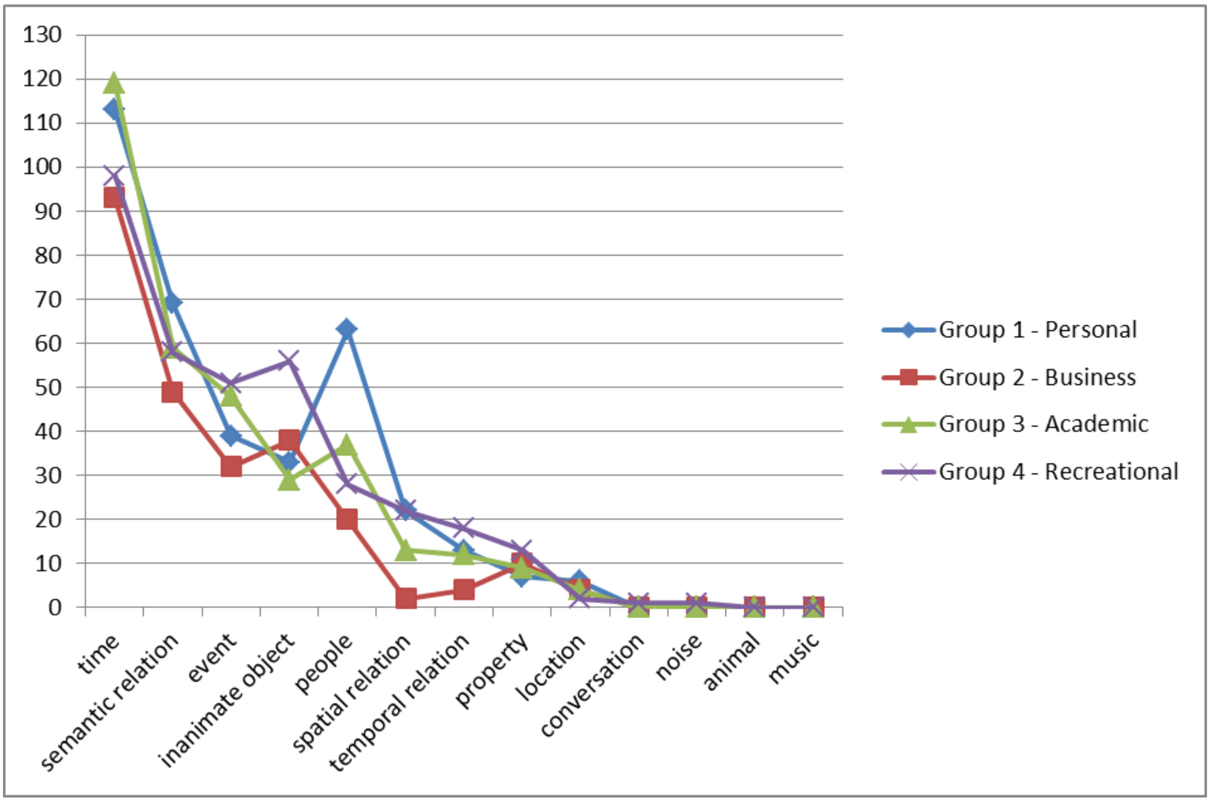


Figure 2.3: Overall usage of tags in MPEG-7 tool per group

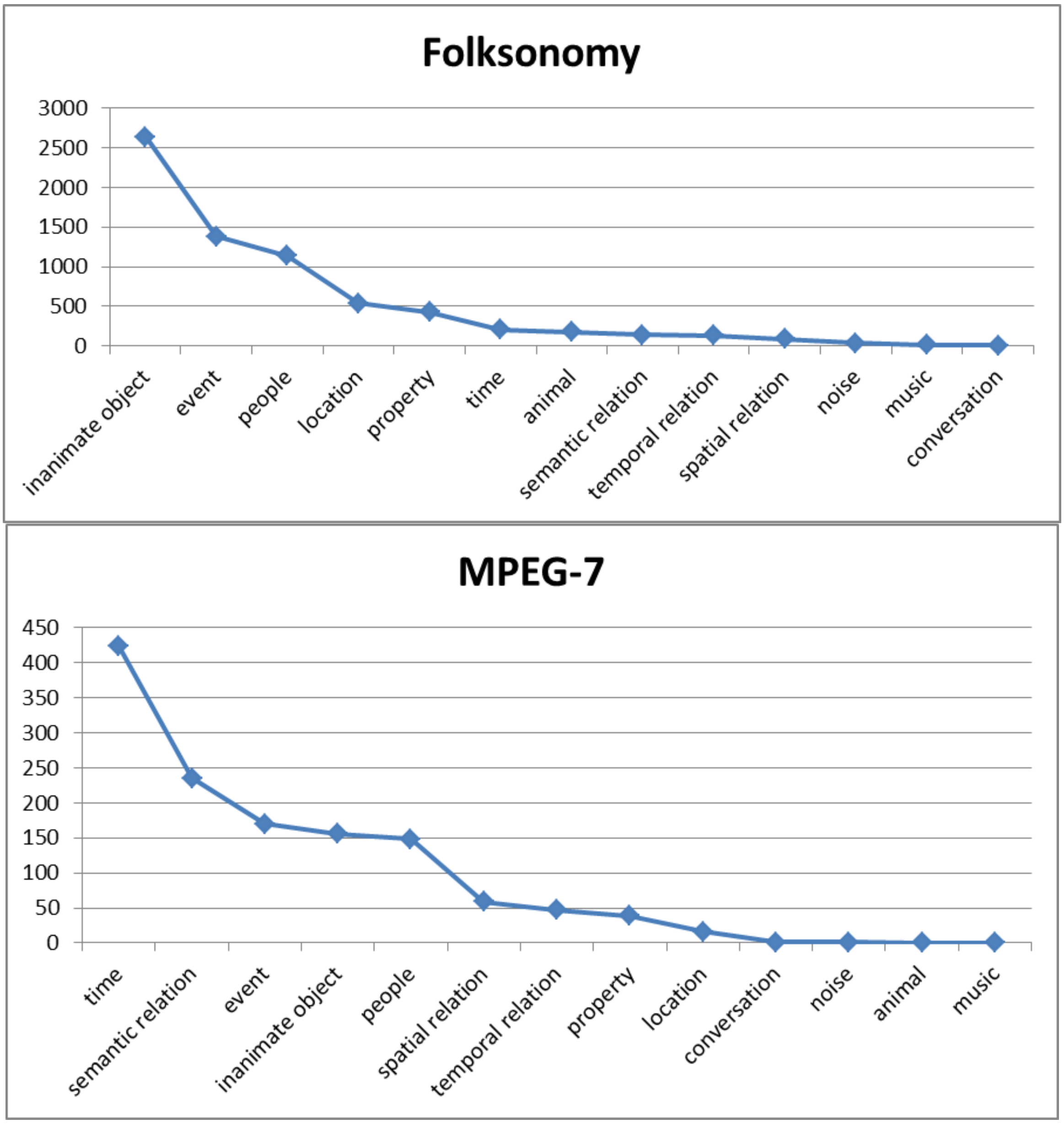


Figure 2.4: Total tag usage in Folksonomy and MPEG-7 tools across all groups

The overall tag usage was then partitioned into most used tags, moderately used tags, least used tags and unused tags for folksonomy and MPEG-7 tools, so that the usage could be further examined. Figure 2.5 presents this tag usage model. The most used tags were the most prominent tags in the memoing stage followed by the moderately used tags. The tags under the least used tags were the ones that were used but their presence was scattered and few during the memoing stage, whereas unused tags were not used at all. The location of a tag in relation to the usage axis depicts whether it was relevant to MPEG-7 tools (above the axis), folksonomy tools (below the axis) or both (on the axis). The intensity of the metadata structure refers to how well-defined the scheme is and thus how the tags are organised and related, if at all. Folksonomy metadata tends to be reasonably unstructured and without a well-defined scheme as users enter tags freely, and thus has low structure, whereas MPEG-7 is very well defined and is thus highly structured. Thus, MPEG-7 benefits from an explicit level of semantic granularity which folksonomy tools lack. MPEG-7's tag labelling was applied to folksonomy data so that the tags' full semantic granularity potential could be exploited. Complexity is how complex the metadata structure is and how detailed the metadata is. The level of relations between tags, the dimensions of the model (one dimensional or multidimensional) and the level of details in the metadata reflect on the complexity of the scheme. Consequently, MPEG-7 is considered to embody high complexity while folksonomies are considered to embody low complexity. The results for each of these four types of tags are considered in turn.

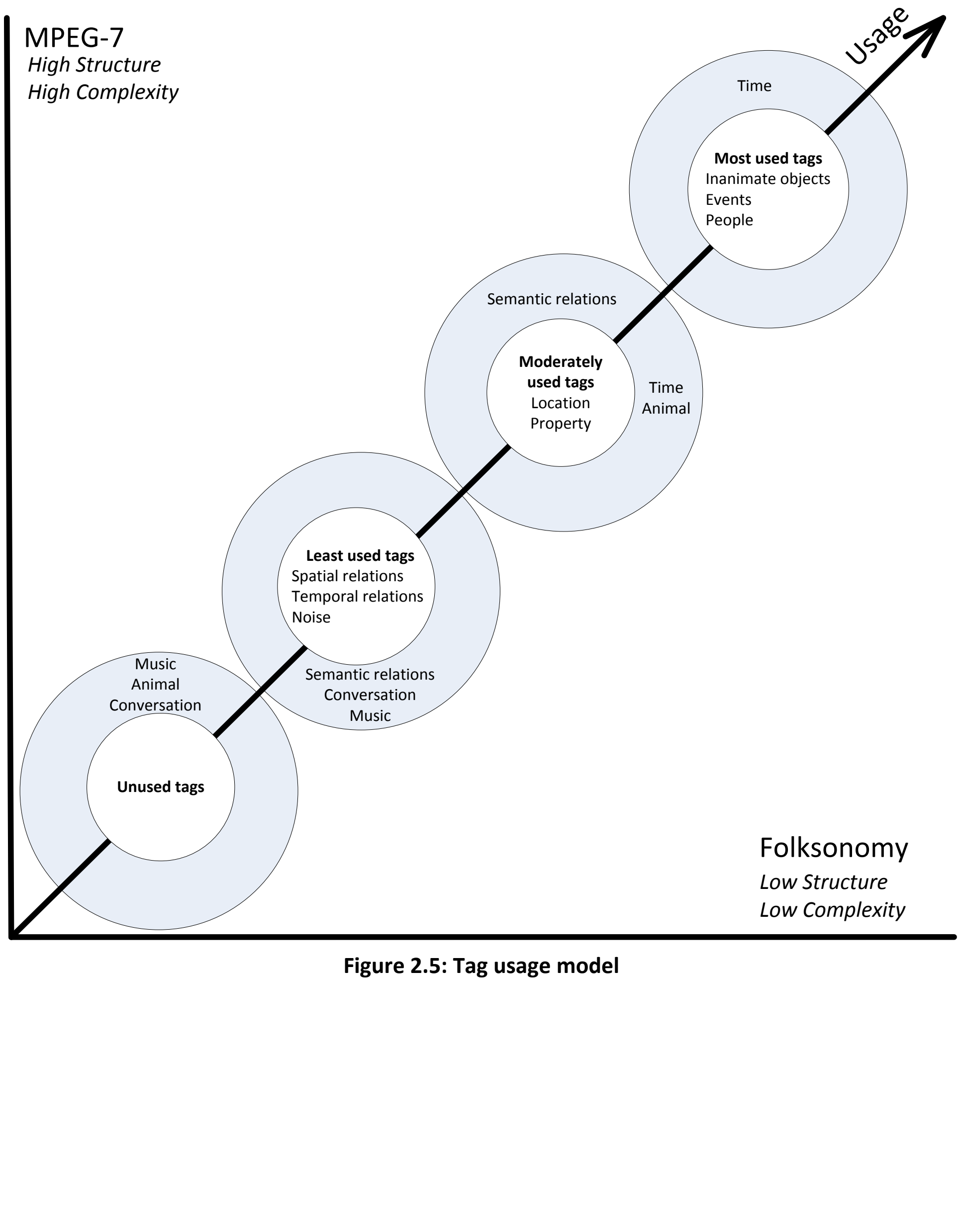


Figure 2.5: Tag usage model

2.2.1.1 Most used tags

According to Li and Lu (2008), recognising the most common tags used by different users when tagging any AV content is valuable for designing an effective collaborative tagging tool. The results indicate that the differences in the use of tags for AV content in different content categories are inconsiderable and, overall, the popularity of tags remains fairly consistent irrespective of these categories. The three tags *Inanimate objects*, *Events* and *People* are the most used tags in both folksonomy tools and the MPEG-7 tool. In the MPEG-7 tool a fourth tag, *Time*, was also apparent as a most used tag. Users also mentioned these four tags as being the most useful tags during their interviews, and suggested them to other users as revealed in the transcripts, both of which corroborate the collected metadata. This suggests that a collaborative multimedia tagging tool should fully support these commonly used tags and prioritise their accessibility. Table 2.1 illustrates the four most used tags during the experiment.

Table 2.1: Most used tags

Tag	Total usage		Average usage		Average deviation		Standard deviation	
	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*
Inanimate Object	2637	156	51.71	3.06	20.567	1.291	25.478	1.815
Event	1384	170	27.14	3.33	8.667	1.320	10.293	1.861
People	1134	148	22.24	2.90	7.670	1.586	9.428	2.174
Time	206	423	4.04	8.30	3.936	2.899	4.919	3.759

Time was the most popular tag employed in the MPEG-7 tool (used 423 times, 9 aupu[16], 2.9 average deviation). In the folksonomy tools, time was moderately used and therefore will be discussed in the next section but for reasons of completeness it is included in Table 2.1 but is shaded. Unlike the folksonomy tools where time was used only moderately, the MPEG-7 tool considers time an integral feature of the media stream and therefore primacy is given to its support during tagging. Consequently, not only are facilities provided for adding one or more time points for each content feature, but users are also prompted to do so. These are therefore the key reasons for the

[16] Average use per user

considerable difference between the usage of time in the folksonomy tools and the MPEG-7 tool. In the interviews and the transcripts, the participants also mentioned that they liked the facility to add time-points to content features and they considered it an important facility in tagging. In the memoing stage, it became apparent early on that the users had used the time point facility extensively: most tags were time stamped with starting point and duration in which they appeared in the video. One user in the recreational group went as far as adding the time points for every single tag he added, while others tended to add time points for more dominant entities.

Inanimate objects were the most popular type of tag used in the folksonomy tools (2,637 total, deviating between 7 and 112 per user, 72 aupu, 20.57 average deviation) and the third most used tag in the MPEG-7 tool (156 total, 4 aupu, 1.29 average deviation). Considering this, it is possible to conclude that not only were inanimate objects the most used type of tags overall but were a popular tag among individual users as well. This is no doubt due to the vast number of inanimate objects reflected within AV content which users wished to give primacy to within the metadata and this was corroborated during the interviews. Throughout the memoing stage, inanimate objects had a dominant presence in the metadata, yet users' attention to detail was somewhat dissimilar. In one of the recreational videos, about cooking fish and chips, all users tagged the inanimate objects 'fish', 'potatoes' and 'chips', but only a few users tagged the inanimate objects 'egg' and 'mixer' that were shown in the making of the batter for the fish. In the same video, there were other users who considered the fish as an animal in the folksonomy tools.

Events were the second most popular type of tag used within both the folksonomy tools (1,384 total, deviating between 8 and 49 per user, 28 aupu, 8.67 average deviation) and the MPEG-7 tool (170 total, 4 aupu, 1.32 average deviation). In the MPEG-7 tool, events were used somewhat more than inanimate objects. The slightly higher prominence of event tags in the MPEG-7 tool may be because of their temporal quality and the support afforded this by COSMOSIS. The fact that any type of occurrence within the AV content can be considered to be an event means that events are highly likely to be identified and thus tagged by users. The high prominence of events also means that they are generally considered by users to be important content, which was also confirmed by the transcripts. A very good example of this is one user in the experiment who added an event called NO-EVENT in both folksonomy and MPEG-7 tools when the video was somewhat uneventful (no events happening and all the objects in the video were motionless), to specifically indicate that there was no events for this segment of the video.

People tags were also used extensively by the participants, ranking third overall in the folksonomy tools (1,134 total, 23 aupu, 7.67 average deviation) and fourth in the MPEG-7 tool (148 total, 3 aupu, 1.59 average deviation). While all content categories featured a significant number of people within the video and images, people tags were used most within the Personal category since the content included users and people known to them, such as their friends and family.

2.2.1.2 Moderately used tags

These tags were not used as significantly to be amongst the most used tags since a significant drop in usage was observed, but there were still ample instances of use of these types of tags for them to be considered important. These tags were also mentioned frequently in the interviews as being important and useful tags. These tags are presented in Table 2.2. The location and property tags were moderately used tags across both types of tools. In the folksonomy tools, time and animal tags were also among the moderately used tags and in the MPEG-7 tool semantic relations were moderately used.

Table 2.2: Moderately used tags

Tag	Total usage		Average usage		Average deviation		Standard deviation	
	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*
Location	535	26	10.49	0.51	3.439	0.719	4.557	0.857
Property	430	27	8.43	0.53	4.787	0.378	5.883	0.513
Time	206	423	4.04	8.30	3.936	2.899	4.919	3.759
Animal	174	0	3.41	0	3.301	0	4.196	0
Semantic relations	139	284	2.73	4.53	0.179	1.697	0.235	1.918

In the folksonomy tools, location (535 total, 11 aupu, 3.44 average deviation) was a moderately used tag but its usage statistics demonstrate its usefulness despite it having significantly less usage than the most used tags. This is likely due to the fact that most AV streams depict content occurring in one or more locations, e.g. even a short 3 minute personal video could depict many different rooms within a house and therefore necessitate several location tags. In the MPEG-7 tools, location (26 total, 0.51 aupu, 0.71 average deviation) was used considerably less, yet this can be explained by the fact that location in COSMOSIS was implemented using semantic relation – locationOf. The transcripts showed that the users were annoyed and confused by this and the users mentioned in the interviews that the facility for adding location was difficult and confusing for them at most times, making it less likely to be used. In addition, the content of still images usually contains only one location and most of the frames of the video will have one location as well. Within a video a single location (e.g. a room such as a kitchen) could be the location of the video for a considerable duration of the video. This means that the participants do not have many locations to tag and therefore location is a moderately used tag. While locations could include other locations that are implicit within the content, users generally did not opt to do this although there were some isolated instances were implicit locations were recognised. For example, in a video featuring a football stadium, one participant tagged the stadium (explicitly depicted in the content) and also tagged England as the location of the stadium (not explicitly depicted in the content).

Property tags were found to normally accompany the most used tags. While property tags in both the folksonomy tools (430 total, 9 aupu, 4.79 average deviation) and the MPEG- tool (27 total, 0.53 aupu. 0.38 average deviation) were not used as significantly as other tags, their usage is still considerable, especially given that property tags cannot be used on their own as they are used to further describe other content features, such as inanimate objects (e.g. their colour) or events (e.g. their status). Due to this associative use which is incompatible with the folksonomic approach, properties are generally not well supported within folksonomy tools, but they are still apparent in the metadata created by users. Folksonomy tools only support free-text tagging and therefore do not provide any specific facilities for adding properties. The data revealed that, in the folksonomy tools, property tags were generally used by users to further describe the most common tags identified previously. As can been seen in Figure 2.6, inanimate object tags were the most likely to be accompanied by property descriptions (41%), followed by event tags (40%) and people tags (38%). Location (26%), animal (18%) and time (17%) tags had associated property descriptions significantly less of the time. In the case of the time tags having associated property descriptions, this was for semantic time only, since no facility existed for associating precise clock date-time points with tags. Hence, for tags such as 'sunny day', 'day' was considered a semantic time tag while 'sunny' was considered as a property of the semantic time tag. As was the case with the folksonomy tools, the most commonly used tags in MPEG-7 tools had high associated usage with properties. People were the most likely to be accompanied by property descriptions (28%), followed by inanimate objects (25%), then events (22%) and finally time (10%), where one or more sets of time points within the media stream were associated with the property. COSMOSIS does not provide a separate facility for animals, users must add animals using the object or agentObject tags. This results in no properties being directly associated with animals in the MPEG-7 tool. Also, COSMOSIS inhibited users from associating properties with locations as it supports locations through semantic relations. The MPEG-7 standard provides different tools for adding location; semantic relations – locationOf and the SemanticPlace DS. COSMOSIS had chosen to support adding locations through the semantic relation but the results show that as users would like to associate properties with locations, it is better to also support location tags through SemanticPlace DS as it allows for adding properties to the location tag. Overall, the results suggest the importance of collaborative multimedia tagging tools providing more effective functionality when adding properties to tags, especially those that are highly associated with properties.

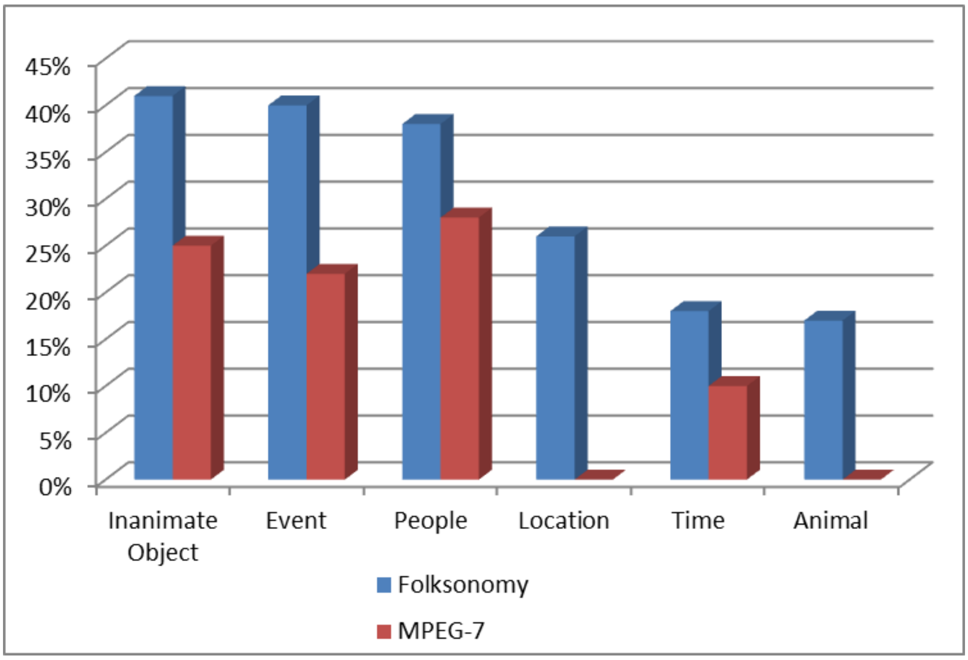


Figure 2.6: Percentage used with property

While time tags were one of the most used tags in the MPEG-7 tool, they were moderately used in the folksonomy tools (206 total, 1 aupu, 3.94 average deviation). There is a very clear reason for this relatively lower usage: folksonomy tools do not provide any means for specifying time-points of different content features, only semantic or precise clock-date time tags can be used. Most instances of time that were found to be used in the folksonomy tools were semantic time such as *morning*, *middle-ages* and *last year* rather than precise clock date-times, yet precise clock date-times, e.g. 12:45 am, were not completely non-existent and a few instances were found. However, semantic time does not have a rich presence within a single image as it normally does not feature more than one semantic time. Similarly, videos (especially short videos as used in this experiment) do not present many instances of semantic time. The fact that, despite this, time was still moderately used by users during the experiment is an indicator of the importance of time when tagging AV content.

Animal tags (174 total, 4 aupu, 3.30 average deviation) were moderately used in the folksonomy tools, which is explained both by the relatively low occurrence of animal content compared to other types of content and the relative importance that users attach to animal tagging. In the MPEG-7 tools, the animal tag was unused by the participants and this will be discussed further in section

2.2.1.4. This is due to the fact that COSMOSIS did not provide any facility for defining animals as an independent entity. Within the experiment groups, the academic group tagged the most animals, especially while using Flickr to tag images, due to the fact that the images given to the participants were animal-rich.

Semantic relations (284 total, 6 aupu, 1.73 average deviation) were moderately used in the MPEG-7 tool while they were least used tags in the folksonomy tools. The semantic relations are a collection of tags concerned with semantic relationships between other content features. When considered in aggregate, they were the second most used type of tag by users of COSMOSIS, ranking above event but below time tags, and were also the most commonly used relation type. However, given that some of the semantic relations were not used at all (the zero figures in the table below) while others were used more frequently, it may seem unreasonable for the aggregate total to carry the same weight as the other tags ranked as most commonly used. However, their use is significant and therefore they were considered to be moderately used. Table 2.3 presents a more detailed view of the particular semantic relations employed by users during the experiment.

Table 2.3: Detailed semantic relations usage

Relation	Usage		Average		Average deviation		Standard deviation	
	Folksonomy	MPEG-7	Folksonomy	MPEG-7	Folksonomy	MPEG-7	Folksonomy	MPEG-7
specializes	3	1	0.04	0.03	0.075	0.19	0.196	0.14
symbolizes	6	0	0.12	0	0.212	0	0.381	0
user	3	27	0.06	0.53	0.113	0.788	0.311	1.046
userOf	0	46	0	0.90	0	1.061	0	1.253
causer	31	87	0.61	1.71	0.572	1.727	0.634	1.910
CauserOf	0	52	0	1.10	0	0.813	0	1.191

2.2.1.3 Least used tags

Spatial relations, temporal relations and noise were the least used tags among both the folksonomy tools and the MPEG-7 tools. In addition to these, in the folksonomy tools, semantic relations, music and conversation were also among the least used tags. Although these tags were the least used, it is important to identify them since they were still used and thus any multimedia content tagging tool should still support their use. However, it seems sensible that the collaborative system should assign lower priority to these tags in both its metadata scheme and interface. Table 2.4 provides an overview of the least used tags found from the experiment.

Table 2.4: Least used tags

Tag	Total usage		Average usage		Average deviation		Standard deviation	
	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*	Folksonomy	*MPEG-7*
Semantic relations	139	284	2.73	4.53	0.179	1.697	0.235	1.918
Temporal relations	136	47	2.67	0.92	0.242	1.061	0.323	1.253
Spatial relations	91	59	1.79	1.16	0.247	0.813	0.423	1.191
Noise	37	1	0.73	0.02	0.455	0.038	0.532	0.140
Music	15	0	0.31	0	0.495	0	0.782	0
Conversation	5	0	0.10	0	0.176	0	0.301	0

In the folksonomy tools, usage of semantic relations, temporal relations, spatial relations, noise, music and conversation was very low compared to the other tags, but the most used were the various semantic and temporal relations, while noise tags were rarely used at all (the noise most users chose to add to the metadata was the noise from an airplane propeller). While the low use of noise can be considered to be because of the deemed lack of significance of noise by tagging users, the low use of semantic, temporal and spatial relation tags derives from the fact that they tend to be considered complex and are not well understood by users. As was the case with the property tags, they are also associative since they indicate a relationship between other content features. They are thus similarly not well supported within folksonomy tools. Conversation tags also represent a complexity of tagging for users as not only is a lot of text required to tag a conversation, verging on subtitling, but also tagging comprehensively enough to show who is holding the conversation at each point can be difficult to determine and tedious for users.

There were some differences in the use of relations between the different folksonomy tools used in the experiment. While the use of relation tags in del.icio.us and YouTube were very similar, this was not the case for the use of relations in Flickr. In Flickr, participants used spatial relations much more than temporal relations because still images do not present temporal relations as clearly as video.

The most used spatial relations were over, behind, under and infrontOf and spatial relations used 83 times in total in Flickr with an average close to 2 instances of use per user. In YouTube and del.icio.us this is notably different with temporal relations being used much more. Temporal relations were used 136 times in total, with most common being during, endOf, before and begins.

In the MPEG-7 tool, spatial relations were used 59 times in total for this single tool, which is much higher than the figure per tool for the folksonomy tools (30.33) although the average per user is approximately the same. Similarly, temporal relations have slightly higher usage when considered against the average per tool figure for folksonomies. In both cases, this is no doubt due to the facilities provided within COSMOSIS to support the tagging of spatial relations. The remaining tags were used barely or not at all. Noise was used once by the participants. Arguably, this is due to both the nature of the content and the deemed importance of these tags for video content by the users.

2.2.1.4 Unused tags

There were no tags left unused by the users in the folksonomy tools. However, conversation, music and animal were not used at all in COSMOSIS. These unused tags were among the least used and moderately used tags in the folksonomy tools which suggests that they were left unused in the MPEG-7 tools either because there were no facilities for using them or their use was too complex for the users to figure out, e.g. because there was no clear option provided for the user. As noted above, conversation is a complex tag since the user needs to add the text of a conversation and define who the speaker is for each segment of the conversation. It may also be necessary for the users to add time points for each segment of the conversation, making it clear when the conversation is occurring in the stream. Therefore, without such facilities, or at least a free-text option, it is impossible to tag conversations. Tagging music requires similar supporting facilities: to tag music a user may need to enter the name of a song, its composer, its performer, year of release, and so on. User may also need to represent lower level features such as tempo, key, and notes. Dedicated facilities for these tags are not available in COSMOSIS. There are also no facilities to tag animals specifically in COSMOSIS, except through specifying them as objects or agentObjects. Users stated in the interviews and the transcripts that they did not recognise this option or it was too complicated for them and they could not decide on which would be the proper approach. Questions such as “how do you add an animal?” were seen to be raised a few times in the transcripts. Consequently, it seems sensible that these tags should still be supported for specialist usage, albeit assigned a low priority in the metadata scheme and made less complex for users to utilise.

2.2.1.5 Co-relation between tags

Another set of key results from the experiment concerned co-relations between the tags; that is, which tags were used with which other tags most often, e.g. for an object tagged in a scene, co-relations are the other tags that tend to be used in conjunction with it. This is an important factor as it leads to the better support for continuous tagging in both metadata and interface of a collaborative multimedia tagging tool. Figure 2.7 shows the relationships between tags for folksonomy tools as a heat map representing the strength of co-relation between two tags. The lighter the colour, the stronger the co-relation (ranging from black, through to dark red, bright yellow and ultimately white). The heat map depicts all possible co-relations between tags and therefore includes instances where there were no co-relations at all (indicated by black). As users are not able to provide time points in folksonomy tools, the relation between time and other tags for folksonomy tools during the experiment was considerably low. Conversely, properties were highly associated with other tags since properties were used to add further detail to, and thus ‘better describe’, these associated tags. There were also strong co-relations between the most used tags, explained by the fact that the users were tagging people, the inanimate objects surrounding them, and the events involving them. Overall, the most common relationships between tags for folksonomy tools discovered from the experiment data were as follows:

Inanimate Object – Property	Inanimate Object – People
Event – Property	Event – Inanimate Object
People – Property	People – Event

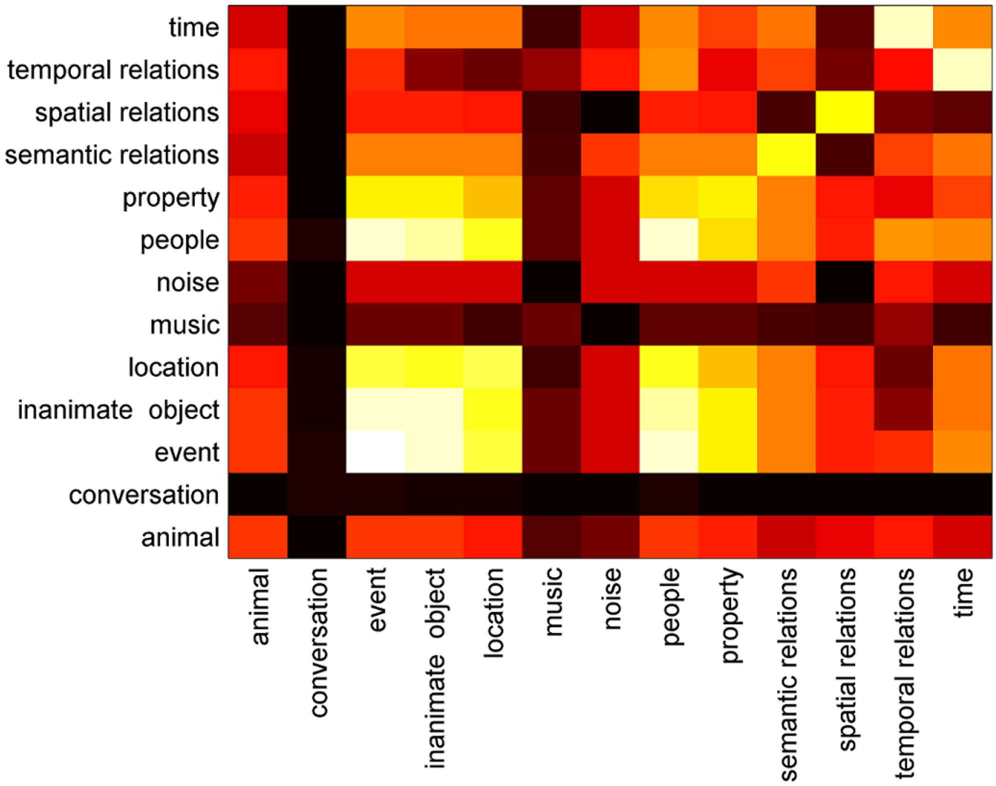


Figure 2.7: Tag co-relations in the folksonomy tools

The heat map diagram in Figure 2.8 shows the relationships between tags for the MPEG-7 tool. As mentioned before the lighter colours represent stronger co-relations between the two tags. As users are able to provide time points here, which was not the case with the folksonomy tools, the relation between time and other tags is extremely high. As noted previously, the use of property in MPEG-7 was lower than in folksonomy tools and this is also reflected among the co-relations as well. Overall, the most common relationships between tags in the MPEG-7 tool discovered from the experiment data were as follows:

Inanimate Object – Time	Inanimate Object – People
Event – Time	Event – Inanimate Object
People – Time	People – Event

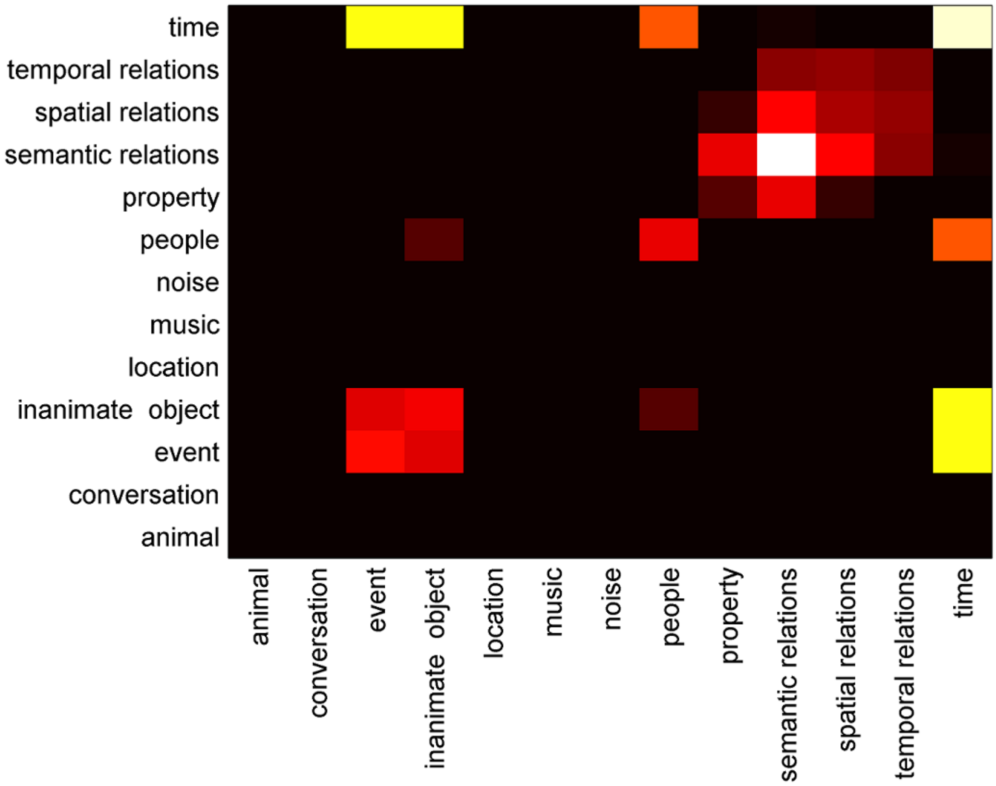


Figure 2.8: Tag co-relations in the MPEG-7 tool

2.2.2 Context

This section discusses the participants' views of their experiences with the tagging context of the tools used in the experiment, which impacts on how efficiently and thoroughly the tools could be used for collaborative tagging. Figure 2.9 summarises the key results, derived predominantly from the interviews and collaborative transcripts while influenced and validated by the metadata.

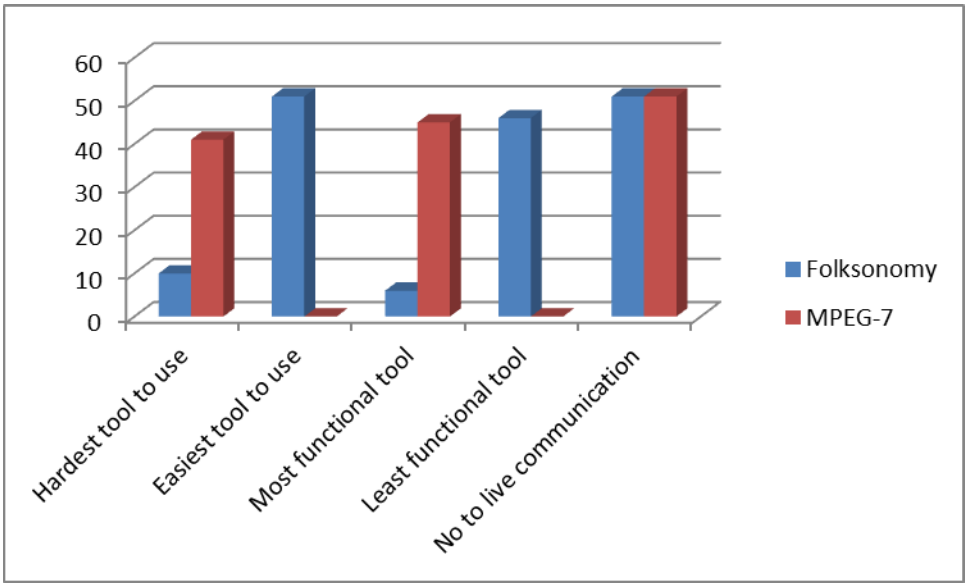


Figure 2.9: Context of the tagging systems

Generally, participants of this experiment found the folksonomy tools easy to use, and in fact all users found the folksonomy tools easier to use than the MPEG-7 tool. Of the 51 users, 36 users (71%) believed that YouTube was the easiest tool to use, while the remaining 15 (29%) preferred Flickr. No users found del.icio.us to be the easiest to use and this was mentioned repeatedly in the transcripts too. Users stated that as the content itself was not present while they are tagging it, it made the process complicated for them. As one user stated, "It is hard to remember how many fish were in the scene when you constantly needed to flick back and forth to check if what you were tagging is truly what was in the video."

Folksonomy tools have very unstructured tagging facilities that normally incorporate free-text boxes with no guidelines or limitations as to what users can or should include, leaving users to freely input as and how they deem fit. Tagging is therefore a very simple procedure which impacts on ease of use considerably. However, as a result, the resultant tags also tend to be simplistic and vague, as they do not overcome possible double meanings or ambiguities, and with content omissions, making the tags lack comprehensiveness. In one case, a user had entered a tag 'key' which made little sense without reading the surrounding tags and watching the video, since it could an object or a property (an adjective such as 'key factor'). Another impact of enforcing only free text tagging with no limitations is that there are no predefined options to choose from which could guide the users towards more understandable, consistent and effective tags. Lack of structure in the tagging process also means there are no facilities for more complicated tags such as semantic, temporal and spatial relations and also no definition and guidance on their use either. Such tags help to shape the metadata scheme into an understandable form as they connect tags together through key relationships. As mentioned previously, the folksonomy tools also do not provide any facilities for adding time points. This reduces clarity of tag understanding as tags become detached from specific temporal segments in the media and apply generally (to the entire media), leaving users unclear when a certain content feature occurs. One facility that most users requested was to have a ‘pool’ of all the tags added for the same content by different users and being able to know which tag was added by which user. None of the folksonomy tools embodied such functionality.

On the other hand, participants found the MPEG-7 tool very hard to use, with 41 users (80%) finding it the hardest of all the tools to use. This is also confirmed by the gap between folksonomy and MPEG-7 tools in the tag usage figures. COSMOSIS utilises a fixed MPEG-7 schema and, unlike the folksonomy tools, free text tagging is not supported. While MPEG-7 does provide facilities for free text tagging within the Multimedia Description Schemes, these are not supported in COSMOSIS since it is intended to be very structured in how it allows users to tag. Potential tags are drawn from a restricted, core set of tags, which are objects, events, agentObjects, temporal relations, spatial relations and semantic relations. Anything that the user intends to tag is assumed to be encompassed by this. As a result of this structure, there are clear and precise definitions for semantic, spatial and temporal relations and users are encouraged to use these relations to build rich content metadata. However, these additional structures make the tagging process considerably more complicated. As mentioned before, these structures created confusion for the users while adding animals and also it made the process of adding locations, properties and conversations more complicated for the users. The user needs to follow the exact pre-defined procedure to be able to add their desired tag. At the same time, however, this structure ensures that the tagging process results in very comprehensive and clear tags. There was no ambiguity in interpreting the tags as they were accompanied by relations and structures that made the intended meaning very clear. The support for adding time points in the form of multiple interface elements to prompt users while they were tagging content features also made the MPEG-7 tool very difficult to use, although once again, the tags generated were the most comprehensive as a result of this because the tags could be related directly to where they occurred in the media stream. As a result, 45 users (88%) stated that COSMOSIS was the most functional tool used within the experiment. Similar to folksonomy tools, the users stated that they would like to have the facility to view all the tags added for a video and be able to see which tags were added, updated or removed by which users.

Communication is crucial to achieving an effective community in collaborative tagging. In the folksonomy tools, users could communicate with each other through forums and taking part in the on-going discussions or by creating a new topic themselves and continuing from there. They were also able to comment directly on the AV content through the tools. Both options were used extensively by the users. On the other hand, as mentioned previously, COSMOSIS is an individual context tool; therefore it does not provide any communication facilities such as forums or comments. Participants were also highly encouraged to use instant messengers to communicate about the tasks they were undertaking using both the folksonomy tools and the MPEG-7 tool. All of the users reported finding this live chat distracting and annoying, even when using COSMOSIS, in spite of there being no other means of communication or collaboration available. One user stated: “It is extremely hard to concentrate on the task at hand when you’re constantly interrupted but what other people are thinking to do”. All users held the view that using the already built-in (asynchronous) functionalities of forums and commenting in the folksonomy tools better supported collaboration, whereas real-time communication only hindered them.

Having presented in detail the results of the experiment, the next section presents the grounded theory that was derived from the usage data as a fuzzy cognitive map. Through the presentation of the grounded theory, implications for collaborative multimedia tagging tools are discussed.

2.3 A Fuzzy Cognitive Model of Conceptual User Behaviour

A fuzzy cognitive map is a cognitive map within which the relations between the elements of a "mental landscape" can be used to compute the “strength of impact” of these elements. Fuzzy cognitive maps are signed fuzzy digraphs (Dickerson and Kosko, 1993; Kosko, 1986; Kosko, 1997). While originally proven valuable in the social sciences, they now have a much wider application in areas as diverse as product planning, economic game theory, and robotics. Fuzzy cognitive maps have the advantage of being able to usefully depict complex relationships between components and ultimate causal reasoning. This makes them an ideal candidate for representing a grounded theory. Figure 2.10 illustrates the grounded theory resulting from the analysis of the experiment data in the form of a fuzzy cognitive map. This represents the conceptual model of user behaviour, and thus it is termed a *fuzzy cognitive model*. It is divided into tagging and context, containing three planes: behaviour, outcome and implication. The behaviour plane represents the behaviour that users demonstrated during the experiment, which was discussed in detail in the previous section. The outcome plane presents the key outcomes that were derived from the grounded theory processes of coding, memoing and visualisation. The implication plane illustrates the design implications that the outcomes imply for a collaborative multimedia tagging system. The directed arcs connecting the nodes represent the positive or negative effect of different elements on each other. This fuzzy cognitive model fills a gap for mapping user tagging behaviour and provides a new insight for researchers into understanding how collaborative user behaviour unfolds.

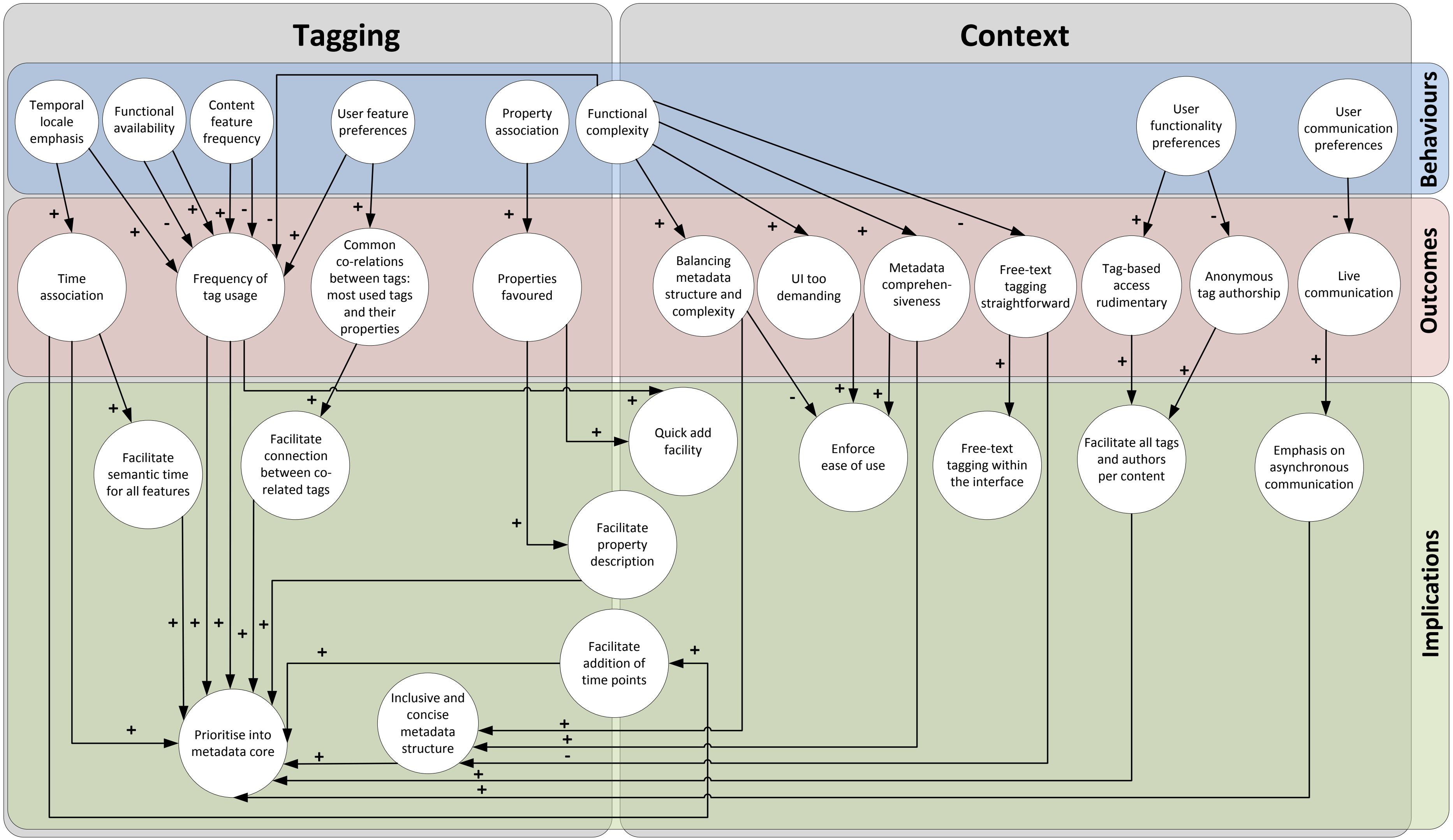


Figure 2.10: A fuzzy cognitive model of user behaviour

2.3.1 Temporal locale emphasis

The experiment revealed a number of significant user behaviours. One key behaviour exhibited by users when tagging metadata was a *temporal locale emphasis* whereby users expressed a high preference towards describing the temporal locale of the content and subsequently representing this within the metadata. As mentioned previously, users like to add time-points and tag semantic time when the facilities are provided for them. As a consequence, this was seen to generate a positive *time association* outcome in the experiment such that an importance was attached to time points by users who frequently added multiple time points per content feature. In terms of tagging, this implies a need to *facilitate semantic time for all features* so that it efficiently and effectively supports the user and, in terms of both tagging and context, it implies the need to *facilitate addition of time points.* Due to their importance, there is a need to *prioritise into metadata core* both semantic time and the addition of time points.

2.3.2 Functional availability, Content feature frequency and User feature preferences

Another significant behaviour was that of *functional availability* where users were found to use tags more frequently when there dedicated functions available for these tags. *Content feature frequency* is another significant user behaviour and it is the frequency in which a content feature is included in the content. As mentioned before some features are included more than others in the content. Another significant behaviour was that of *user feature preference* and it represents which content features are preferable by users for tagging. *Functional availability* and *content feature frequency* have both positive and negative effect on the *frequency of tag usage* outcome, while *temporal locale emphasis* and *user feature preference* have only positive effect on it. The outcome *frequency of tag usage* encapsulates the most used tags, the moderately used tags and the least/not used tags. High *content feature frequency* along with high *user feature preference* results in the most used tags. As the most used tags in the folksonomy tools and the MPEG-7 tool, for any metadata scheme there us a need to *prioritise into metadata core* the tags time, inanimate objects, people and events. The tags inanimate objects, people and event were the most used tags in both tools, but time is also included as a most used tags (in folksonomy tools time was moderately used and thus not so far removed from the most used tags) as the interview transcripts showed that the users considered time to be an important feature of multimedia content. To hasten adding of these frequently used tags, the support of a *quick add facility* would enable the full process of tagging to be circumvented. For example, guidance and structure would not be enforced in this mode, enabling advanced users to tag more speedily without restriction.

The moderately used tags, namely location, property, animal, and the semantic, spatial and temporal relations, they have been deemed sufficiently significant by users and therefore create a need to *prioritise into metadata core*. Relations in particular connect different elements of the metadata scheme together, lending greater meaning and structure to the metadata created by the users, and the most common relations, which were found to be the semantic and temporal relations, should be given key support within the metadata scheme and the system interface, e.g. through pre-selection. However, the noise and music tags (least/not used tags) were used so little that they should not be supported specifically so as not to impinge upon the efficiency and effectiveness of the underlying metadata scheme and the subsequent system interface through redundancy. One option is to allow users to represent such features using other tags that are supported, e.g. noise and music could be represented as events.

2.3.3 Property association

Our results show a trend in user behaviour with regards to *property association*, meaning that users tend to associate tags with additional properties, specifically the most used tags which tended to be associated with properties nearly 50% of the time. This has a positive impact on the *properties favoured* outcome and therefore the *facilitate property description* should be provided in both tagging and context. The metadata scheme should provide the means for supporting properties for different content features and therefore there is a need to *prioritise into metadata core* for properties, while the interface should facilitate the tagging process through prompting (particularly prompting after a most commonly used tag is added), which has proven to be valuable in the case of the MPEG-7 tool used in the experiment, leading to resultant metadata that is much richer.

2.3.4 User feature preferences,

The user behaviour of *user feature preferences* also effects the *common co-relations between tags* outcome. As users tend to add certain tags after they have added a tag, the *common co-relation between tags*, specifically the co-relation amongst the most used tags and between the most used tags and their properties, is a significant outcome of the experiment. This leads to the implications that the tagging should provide efficient facilities for connection between co-related tags while the context should have facilities to simplify adding co-related tags in the interface.

2.3.5 User communication preferences

The significant user behaviour of *user communication preferences* represents users' preferences for the type of communication facilities they would like to have while tagging. Since all participants acknowledged the importance of communication but felt live, synchronous communication was inappropriate and hampered their collaborative efforts, emphasis within the system should be placed on asynchronous communication channels, such as the forums and comments used in the folksonomy tools. The core metadata scheme should provide support for this asynchronous communication and therefore *prioritise into metadata core* for this type of communication. Also the context should provide facilities in the interface with *emphasis on asynchronous communication*.

2.3.6 Functional complexity

Functional complexity is another significant user behaviour and it represents the complexities and confusions arising from the tagging and context of a system. Given the contrasts between the conceptual models for the folksonomy and MPEG-7 tools, there is a need to minimise complexity of the tagging and the system interface, while maintaining rich metadata schemes and ensuring the system is sufficiently easy to use. The MPEG-7 tool's interface and metadata were deemed too hard to use, stating tagging and interface complexity was undesirable, while the comprehensiveness of the metadata was appreciated which impacts upon: the *UI too demanding* outcome, as the more comprehensive the metadata gets, there are more options and details for it in the interface which makes it difficult to use; the *balancing metadata structure and complexity* outcome, because the more detailed the structure gets, the more complex it will be and while the users found the structure valuable they also felt it was very restrictive; and finally *the metadata comprehensiveness* outcome, as the more comprehensive the metadata gets, there is more detail in the, and volume of, metadata which makes it more useful. Conversely, the users found the folksonomy tools easy to use but the metadata was confusing, impacting the same outcomes. It is necessary to simplify the structure to increase clarity while balancing elements to preserve comprehensiveness, thereby reducing tagging and interface complexity and retaining ease of use. These implies an e*nforce ease of use* implication and impacts upon the *inclusive and concise metadata structure* implication, which in turn has high impact on *prioritising into metadata core*. One immediate way to do this would be to merge the most popular interface components from both types of tools.

Similarly, as part of *functional complexity*, there is a need to impose structure but without losing the advantages of free-text tagging. While the structure and defined process in the MPEG-7 tool helped users form their tags with few omissions, it also made the tagging process more cumbersome and complicated. The *Free-text tagging straightforward* outcome within a structured process is another outcome which balances both elements. This implies that in tagging, the metadata scheme should provide facilities for free-text tagging and context should have *free-text tagging within the interface*.

2.3.7 User functionality preference

Finally, *user functionality preference* is another significant user behaviour that is concerned with the fact that users like to see all tags for a media resource and users would like to know who added each tag, implying a real need for authorship management which leads to the outcomes: *tag-based access is rudimentary* and *anonymous tag authorship*, which is obstructive. Therefore the tagging should support authorship in the metadata core scheme while the context must provide efficient facilities for showing all tags and their authors for the media.

2.4 Summary

The chapter contributes a fuzzy cognitive model of user behaviour that explains relationships between tagging behaviour and context. The model was derived through a user experiment with existing multimedia tagging tools. The data sets collected were analysed using grounded theory and the model constructed was visualised as a fuzzy cognitive map. The next chapter discusses the design and development of the MPEG-7 modelling framework, called MC^2, from the conceptual model of user behaviour developed in this chapter, and the online service implemented as a proof of concept of this framework.

Chapter 3: The MC^2 Framework and Online Service

Tagging of multimedia resources, whether unstructured or structured, is a laborious and time-consuming process especially when carried out by a single tagger. However, such effort can be reduced and the quantity and quality of metadata increased when carried out by a web community. Such collaborative tagging communities can create and maintain better metadata content models for multimedia resources more effectively than single taggers working alone. The aim of this chapter is twofold: first, to design and develop a framework for such collaborative tagging of multimedia content based on the fuzzy cognitive model of user behaviour developed in the previous chapter and, second, to implement an online service from this framework. The framework is aptly called MC^2.

3.1 The MC^2 Framework

The fuzzy cognitive model of user behaviour reveals a set of key implications which are tabulated in Table 3.1. This set shall form the basis for designing and developing MC^2.

3.1.1 Research Approach: UML use cases and Grounded Theory

Use cases are used to specify the design requirements of the framework. The use cases are then mapped to architectural components of the MC^2 framework. In order that the functional system requirements and implications could all directly influence the framework design, a grounded theory approach has been adopted throughout such that all of these influences could emerge into a well-defined and grounded specification that fully capture the design requirements.

First, grounded theory is applied to the set of key implications in Table 3.1 and a set of evident requirements in order to derive a use case diagram that specifies the functional requirements of the framework. Secondly, grounded theory is applied to this derived use case diagram in order to derive the architectural components of the MC^2 framework. Evident requirements are considered those which are not explicitly specified but are evidently necessary in order for the framework to be fully functional. For example, there is a requirement for users to upload media so that it is available within the framework. The experiment participants had no direct input into either stage except through the preliminary experiment results. The outcomes of both stages are presented in the following two sections.

Table 3.2 compares the methodology for each of these two stages. Both approaches follow a similar practice, as they are both an adaptation of grounded theory to achieve a grounded truth. The data input in the two approaches share commonalities, i.e. the implications, due to their importance of establishing the core requirements and the collaborative content modelling framework. In both approaches, the input data will go through a cycle of theoretical and constant comparison to derive the initial, typically high-level, elements of the final diagram or framework. In the memoing stage these are then refined and more precise elements are identified and described together with their inter-relationships (e.g. extend and include in the case of the use case diagram, process and data flows in the case of the framework). These memos are then visualised as grounded theories embodied by a use case diagram and architectural framework diagram respectively. Now each is described in turn.

Table 3.1: Implications revealed in the fuzzy cognitive model of user behaviour

Implication	Elaboration
Facilitate semantic time for all features	Semantic time should be supported and able to be connected to all other features.
Facilitate connection between co-related tags	User behaviour produced co-relation outcome, showing users tend to co-relate certain tags.
Prioritise into metadata core	Certain tags should have higher priority in the metadata based on the user behaviour and preferences reported by the experiment.
Inclusive and concise metadata structure	Comprehensive metadata is appreciated by users but its complexity is not.
Facilitate property description	Users tend to actively associate properties with other content features.
Facilitate addition of time points	Users tend to add time-points for content features.
Enforce ease of use	Complexity and comprehensiveness go hand in hand and thus are related for both interface and metadata.
Quick add facility	Users prefer to add most used tags quickly.
Free-text tagging within the interface	Users appreciate free-text tagging.
Facilitate showing all tags and authors per content	Need for authorship management to counter rudimentary tag-based access and anonymous tag authorship.
Emphasis on asynchronous communication	Users found live communication unproductive.

Table 3.2: Comparison of grounded theory in this two-stage research approach.

	Stage 1: Functional requirements	***Stage 2: MC^2 framework***
Data	Implications and evident requirements	Stage 1 use case diagram (use cases, use case relations, user actions), implications
Open Coding	Initial use cases	Initial subsystems and components
Memoing	Refined use cases, use case relations and user actions	Refined subsystems, components, process and data flows
Grounded Theory	Use case diagram	Collaborative content modelling framework

3.1.2 Functional Requirements

The use case diagram that was produced using the grounded theory approach described above is shown in Figure 3.1. It represents the full extent of user actions when content modelling multimedia collaboratively and the functionality needed to support them.

Table 3.3 summarises how the use cases map to the implications explained previously. It serves to validate the grounded theory approach used here by showing how each implication is addressed by one or more use cases. Use cases which were the result of evident requirements only, such as *manage media*, are excluded from the table since they serve to enable and support other use cases.

As can be seen, the *add/edit tags* use case lies at the core of the requirements and caters for the tagging processes of the framework. Together with its various extensions, it provides the means for tagging the major content features that were revealed during the open coding process, such as *objects* and *events*. It is fully based on prioritising the tagging behaviour of the users and addresses implications. *Add/edit spatiotemporal locales* facilitates both semantic time and time point features as an extension in order to provide inter-referential awareness. *View followup tags* serves to facilitate connection between co-related tags by presenting the user with other tags that they might be interested in adding after a certain tag has been added, thereby improving connectedness.

Several use cases serve to enforce ease of use. *Free text tagging* provides users with the facility to describe content free of structural limitations, while the *view predefined tags* and *predictive tagging* use cases form a recommendation component which checks tags, recommends new tags, suggests correct spellings and word usage, thereby enforcing synonym control and tag expressiveness. *View predefined tags* supports metadata propagation through tag reuse. The *quick add* use case enables users to add the most commonly used tags easier and more swiftly.

The implications also suggest that users exclusively prefer asynchronous communication over live chat. Consequently, the *communicate* use case supports *messaging* and *forum* capabilities and synchronous communication is not supported.

To facilitate the association of tags and authors on a per content basis, to propagate metadata easily, and to support identity and semantic awareness, *profiling* supports all major use cases by recording every action taken by the user in their profile. These actions can be retrieved through *view user profile*, *view media* and *view model* such that users are able to view the profiles of other users, their actions and the tags contributed by them to the content model, while *retrieve media* supports tag-based ranking through ranked searching of the repository.

Finally, the *manage media* use case and its extensions emerged from the evident requirements and serve to populate and manage the media repository. The use of a player component within the framework is supported through the *view media* use case.

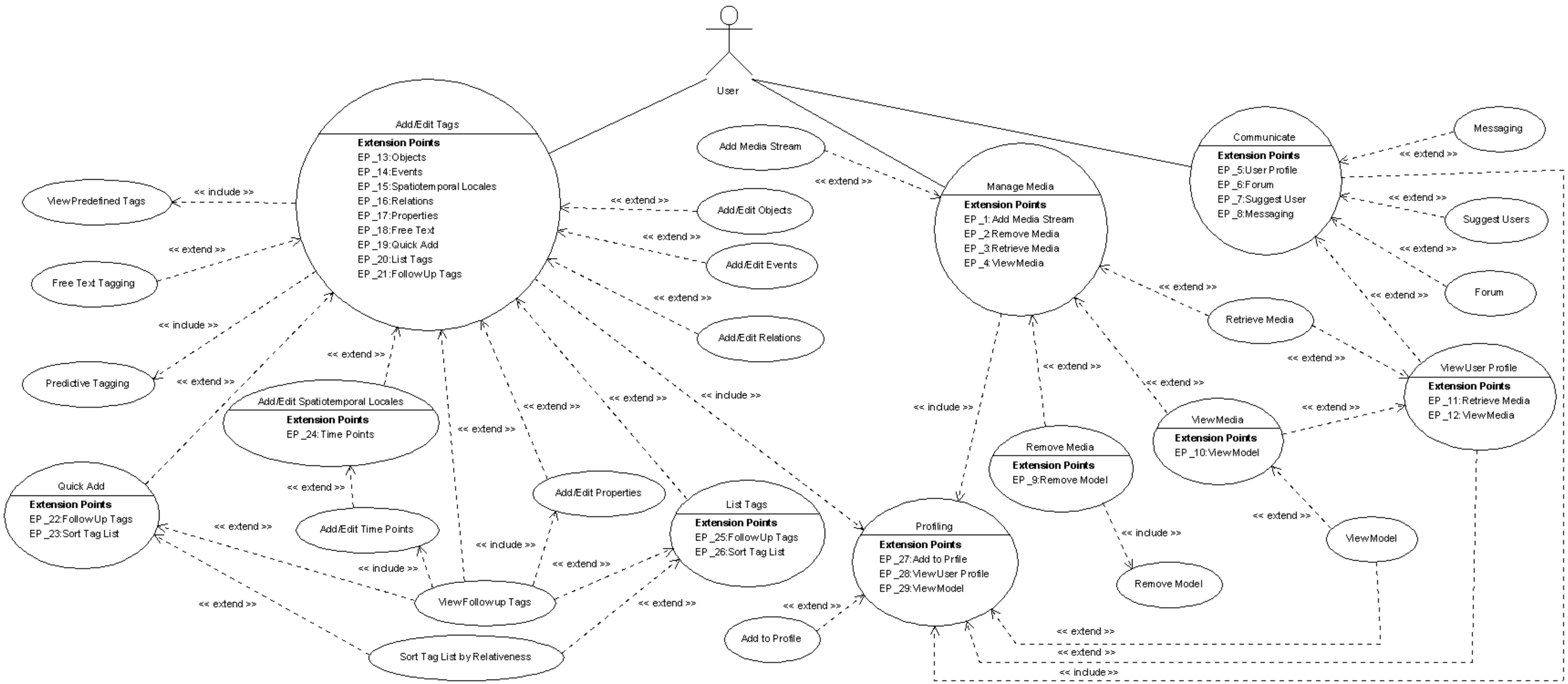


Figure 3.1: Use case diagram portraying the functional requirements for a collaborative content modelling framework.

Table 3.3: Implications and their supporting use cases

Implication	Supporting use cases
Facilitate semantic time for all features	Add/Edit Spatiotemporal Locales
Facilitate connection between co-related tags	View FollowUp Tags, Add/Edit Properties, Sort Tag List by Relativeness
Prioritise into metadata core	Add/Edit Objects, Add/Edit Events, Add Relations, Add/Edit Properties, Add/Edit Spatiotemporal Locales
Inclusive and concise metadata structure	Add/Edit Objects, Add/Edit Events, Add/Edit Relations, Add/Edit Properties, Add/Edit Spatiotemporal Locales, Free Text Tagging
Facilitate property description	View Followup Tags, Add/Edit Properties
Facilitate addition of time points	Add/Edit Time Points
Enforce ease of use	Free Text Tagging, View Predefined Tags, Predictive Tagging, Quick Add, View Followup Tags
Facilitate showing all tags and authors per content	Profiling, Add to Profile, View Model, View User Profile, View Media, Retrieve Media
Emphasis on asynchronous communication	Communicate, Forum, Messaging
Quick add facility	Quick Add
Free-text tagging within the interface	Free Text Tagging

3.1.3 MC^2 Collaborative MPEG-7 Content Modelling Framework

This section presents the MC^2 framework for collaborative multimedia content modelling using the MPEG-7 standard that was derived using the grounded theory approach presented in Section 3.1.1. The framework is given in Figure 3.2 and each subsystem of the framework is described in detail.

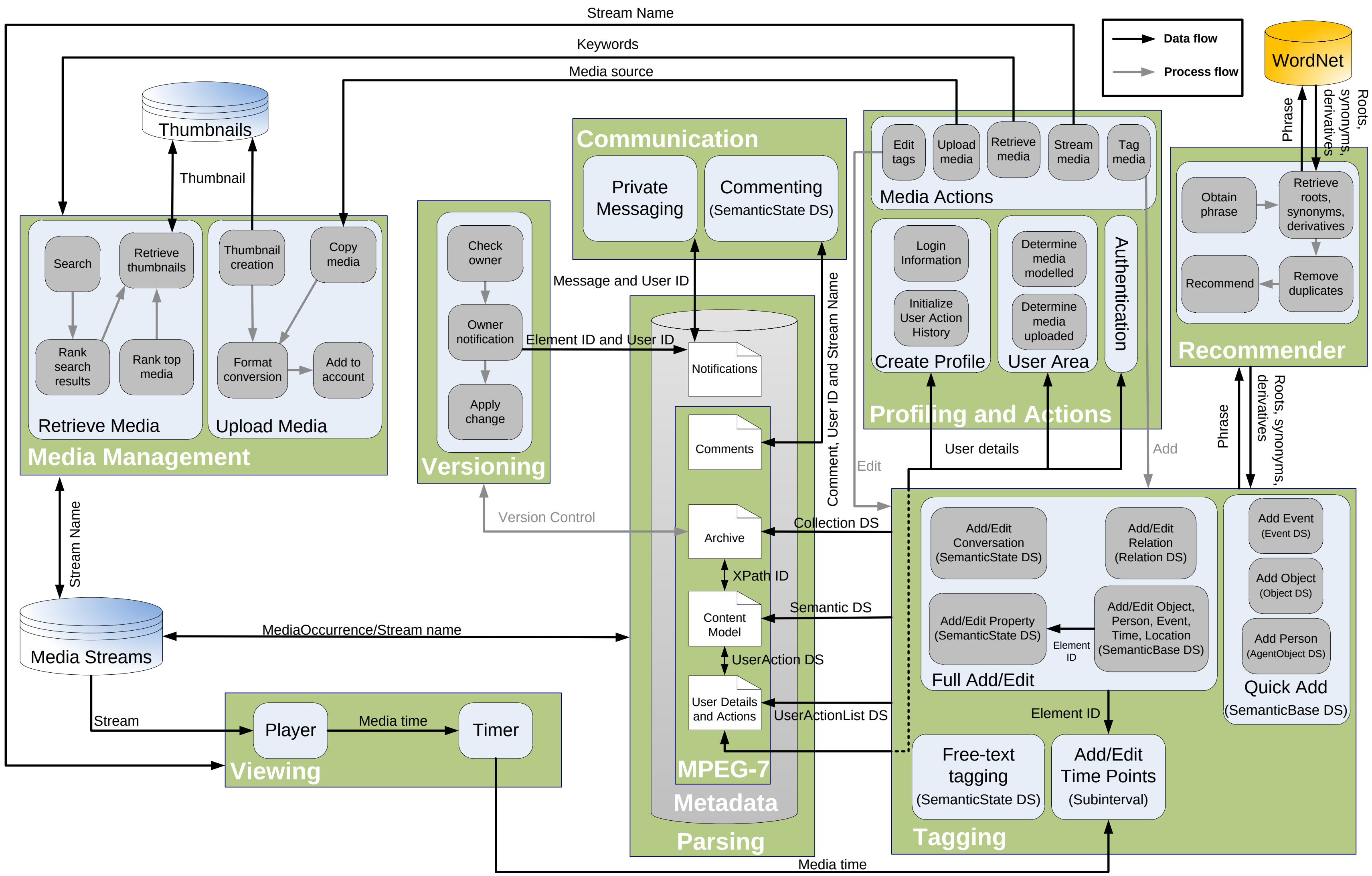


Figure 3.2: MC² collaborative multimedia content modelling framework

3.1.3.1 Parsing and metadata repositories

The *parsing* subsystem lies at the heart of the framework and is responsible for creating, parsing and updating the metadata which is all stored in MPEG-7 XML (ISO/IEC, 2003; ISO/IEC, 2004; ISO/IEC, 2005b). It encompasses five different metadata repositories catering for user details and actions, the content model, an archive for historical changes and user profiling, comments, and notifications. Whereas frameworks commonly use non-XML databases for non-content model data such as user details and comments, MC^2 uses MPEG-7 for all data storage and retrieval purposes throughout the framework, not just the content model. This single MPEG-7 schema provides the advantage of data uniformity throughout the framework.

The *content model* repository represents all media stream metadata. Its schema is defined as a bespoke MPEG-7 profile consisting of various description schemes (DSs), shown in Figure 3.3. MPEG-7 profiles are subsets of the standard that apply to specific application areas (Troncy et al., 2010) and hence the MC^2 profile is a subset catering specifically for the requirements established by this research. To take into account frequency of tag usage while balancing metadata structure and complexity, user feature preferences, content feature frequency and functional availability have been balanced alongside the functional complexity of the content model. To incorporate the tags most used by users (from the *add/edit tags* use case and its extensions), the Semantic DS acts as a foundation for the majority of tags incorporated into the profile. AgentObject DS is used to describe people in different forms as Person DS, PersonGroup DS or Organization DS; Event DS is used for defining events; and Object DS is used to define objects. No specific description scheme exists within MPEG-7 for modelling animals, therefore the profile supports the use of Object DS or AgentObject DS, both of which are equally suitable, depending on the user’s viewpoints and preferences. The SemanticState DS enables modelling of people, event and object properties, while the SemanticTime and SemanticPlace DSs facilitate modelling of spatiotemporal locales for people, events and objectives (semantic time may also be represented per se within the profile). All Semantic DS tags may include MediaOccurrence elements for associating start time and duration.

As with animals, there are no specific noise or music representation structures within MPEG-7; thus, noises are represented through an Event DS inside the Semantic DS, while music is represented via the Object DS. Further, more advanced, music metadata may be described using the CreationInformation DS, which encapsulates the Creation DS and RelatedMaterial DS and may be linked to the Object DS using the Reference type. Conversations are another content feature not specifically supported in MPEG-7. Therefore, they are represented through the SemanticState DS within the profile, which supports time stamping through MediaOccurence, with the text of the conversation stored as an unlimited number of AttributeValuePair elements. Conversations may be linked to AgentObject DSs using the Agent Semantic Relation so that the speaker of segments of conversation can be identified. All semantic, spatial and temporal relations are represented within the Graph DS, nested in the Semantic DS, facilitated by the Relation DS which structures the MPEG-7 profile and inter-relates tags. The UserDescription DS links the user activities to the content model and other interactions taking place in the system.

The *user details and actions* repository represents all metadata with regards to the *profiling* and *view user profile* use cases. All data is stored under the top-level type, UserDescriptionType, which encapsulates the profile and login of the user as User from AgentType and all the user’s actions under the UsageHistory DS. The full structure is summarised in Figure 3.4. UserActionList DS stores information such as action type, action time, stream name and a link to the element in XPath format which are stored respectively in ActionType, ActionTime, ProgramIdentifier and ActionDataItem. To support action types that are specific to this collaborative multimedia content modelling framework, the MPEG-7 profile defines new ActionTypes as follows: Comment, Create, Remove, Update and Upload, which are used to represent the actions of commenting on media, creating a new tag, removing a tag, updating an existing tag, and uploading a new media stream, respectively.

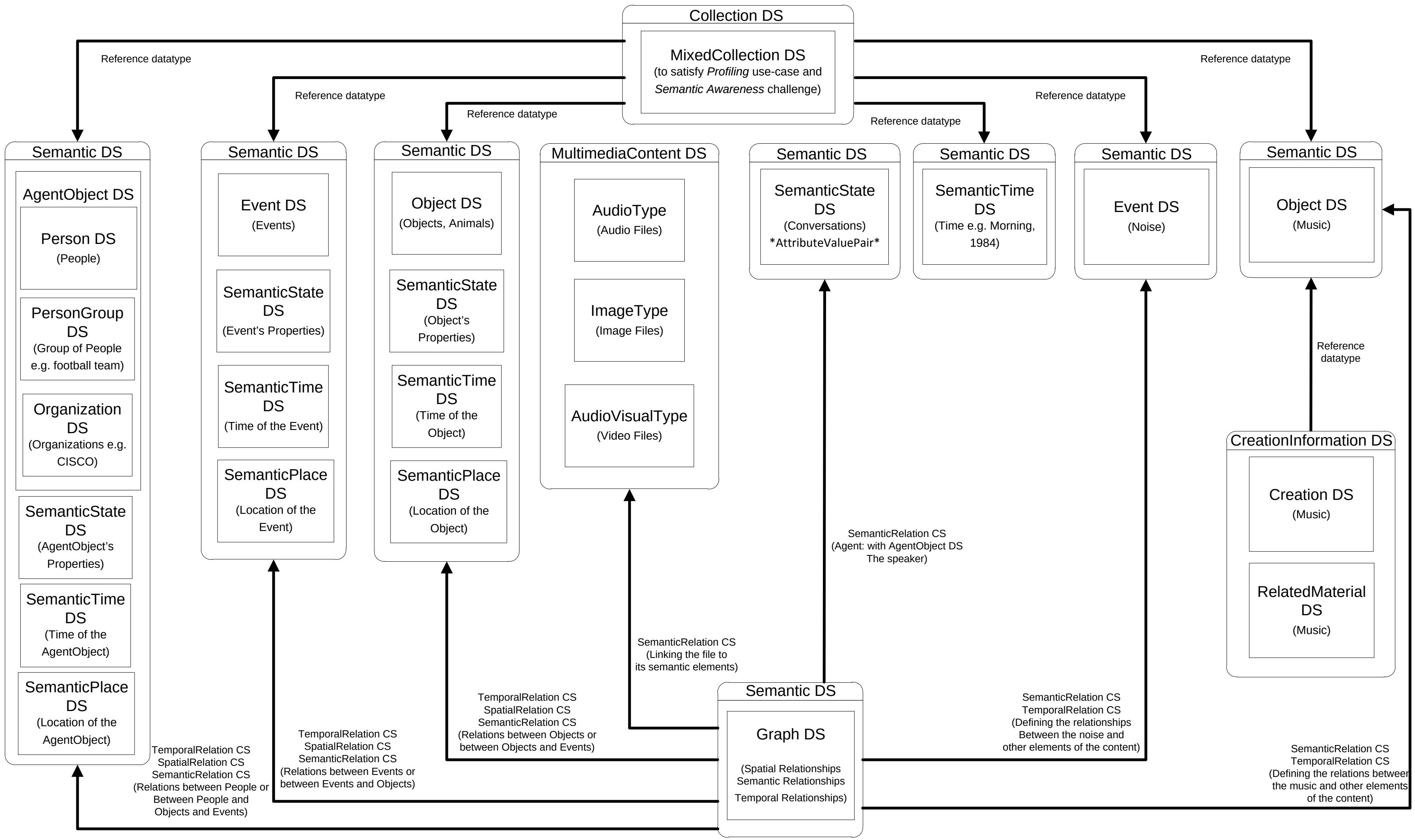


Figure 3.3: MPEG-7 profile for collaborative multimedia content modelling communities

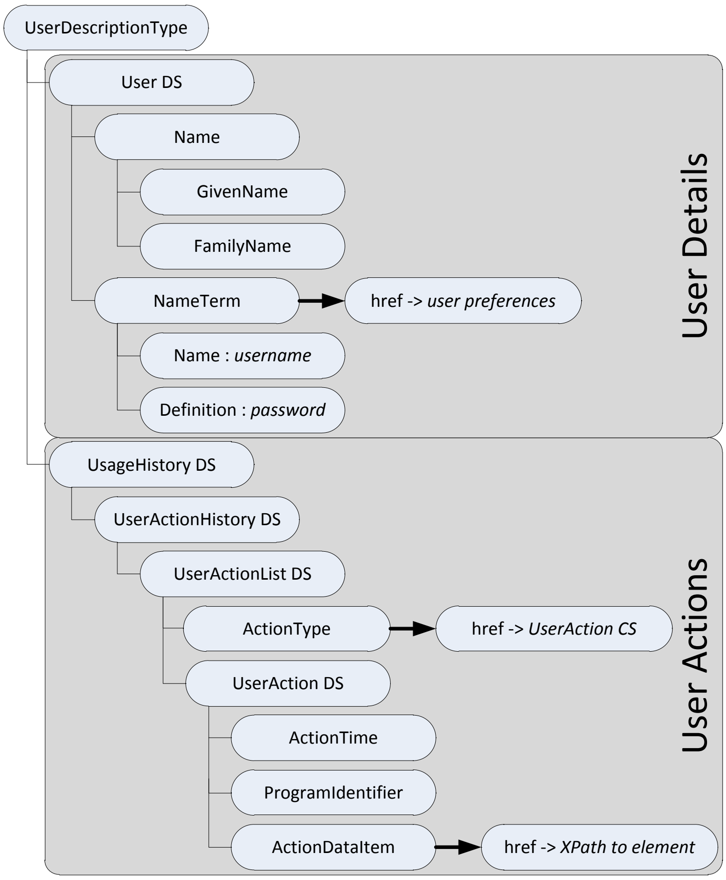


Figure 3.4: Structure of MPEG-7 metadata for user details and actions

The *archive* repository tracks and represents changes to the content model undertaken by users (*profiling* use case). It is represented by Collection DSs, where each Collection DS is from the MixedCollectionType and can contain a SemanticBase DS or Relation DS under the Content element and a reference to the user who has performed an action in the ContentRef element. This structure encapsulates any tag that has been added, removed or updated with a reference to the user who has made the change, therefore representing the full historical evolution of the content model over time, which in turn facilitates version control and rollback operations. The *comments* repository encapsulates the user comments made about each media. It contains a SemanticState DS for each comment that supports a link to the posting user, the title and text of the comment, and a link to the referred to media. Finally, the *notifications* repository contains system notifications and users' private messages to each other, facilitating user messaging.

3.1.3.2 Tagging and recommender

The *tagging* subsystem manages all tagging action. The *recommender* subsystem manages root and synonym recommendations after a tag is added by a user. The *tagging* subsystem results from the *add/edit tags* use case and its extension points and is responsible for adding and editing tags for media streams and storing them in the metadata repositories. Every tag added by this subsystem will add: an element in the Semantic DS to represent the tag in the content model repository, a UserActionList DS in the user details and actions repository to specify which user added which element in the content model, and a Collection DS in the archive repository to represent the element added in the model and a reference to the user.

The *full add/edit* subsystem is responsible for comprehensively adding or editing a tag which conforms to the MPEG-7 profile of the content model. There are four processes that can be chosen based on the type of tag to be added or edited. Conversations are added in the form of a SemanticState DS, consisting of the time, duration and text of the conversation together with a reference to the speaker. Objects, people, events, times and locations are added in the form of a SemanticBase DS set to the corresponding type. Semantic, spatial or temporal relations are added via the Relation DS and have sources and targets that can be selected from tags already present in the content model. Additionally, multiple properties and time points may optionally be added or edited and associated with these tags, whereby the ID of the tag in question is stored as the source of the time point or property (property name and value). Where properties are added to existing content features, the user is required to select the source. In all cases, properties are stored in a SemanticState DS. The (optional) addition of time points enables conversations, properties, relations, objects, people, events, time and location to be associated with particular segments of the media streams, whereby SubInterval elements are added to the MPEG-7 content model containing the start point and (optionally) duration for each tag. Time points are derived from the *viewing* subsystem described below. Where tags are edited (updated or deleted), the user actions are described in a UserActionList in the user details and actions MPEG-7 repository and the tags are modified within the Semantic DS in the content model and replicated in a Collection DS in the Archive repository. The *quick add* subsystem rapidly facilitates adding the most used tags from the experiment, specifically objects, events and people. In this case, only the tag name is stored; though properties and time points may be added later when editing. Through the *free-text tagging* component, users may enter text for their tags freely, e.g. as they would do in YouTube. These free-text tags are stored in a SemanticState DS, as prescribed by the MPEG-7 profile.

Whenever tags are added to the content model, the *recommender* subsystem will retrieve the phrase and check the WordNet dictionary database for its roots, synonyms and derivatives. Then it will remove the duplicates and make a recommendation to the user from the remaining set so that more tags may be added if required, thereby broadening the range of meanings and synonyms. This subsystem maps to the *predictive tagging* and *view predefined tags* use cases described previously.

3.1.3.3 Versioning and viewing

Derived from the *profiling* use case, the *versioning* subsystem manages revisions made to content models. When the content model is updated, the *check owner* process is instantiated which identifies the author of either the corresponding section of the content model for an update or the user who originally uploaded the media stream if a new tag is being added. In cases where the same user is not performing the action, then the author or uploader (whomever is appropriate) is notified through the *owner notification* process. Then, the *apply change* process makes the change effective in the system if the owner accepts the change or it will rollback the content model to its prior state if the owner rejects it.

Derived from the *view file* use case, the *viewing* subsystem manages all stream playing functionality. The *player* receives the media stream from the *media stream* repository and streams it to the user. Having the stream active while tagging helps improve tagging accuracy and the preliminary experiment revealed that users tend to use time points if facilities to support this are provided. Thus the provision of a *timer* to handle all timing requests from the *tagging* subsystem serves to encourage and support this with a view to enabling comprehensive and accurate content models.

3.1.3.4 Profiling and Communication

The *profiling and actions* subsystem is derived from several use cases, namely *profiling*, *view user profile* and *add to profile*. It is manages all user actions and activities. New users join the system through the *create profile* component which stores *login information* in the user details and actions repository after which the *initialise user action history* process will create a new (empty) action history for the user. To support personal user areas within the utilising system, the *user area* component enables reviewing of media uploaded to the system and media content modelled to date and commented upon. In this way, users are able to efficiently and effectively track their own behaviour and history within the system. The *media actions* component incorporates various processes that allow users to perform different actions on the media, namely uploading, retrieving, tagging and streaming media, as well as editing tags.

The *communication* subsystem manages user communication facilities in the system. The *private messaging* component handles the creation and reading of private messages between users. It stores the messages in the system by storing the IDs of the sender, receiver and the message, so that they may be retrieved on request by the user. The *commenting* component caters for media comments, storing the user ID, stream name and comment text and handling retrieval of comments. Commenting has been ascertained as being a well-established and desired functionality of any collaborative community, according to both the experiment results and the research literature described previously.

3.1.3.5 Media management

This subsystem caters for the uploading and retrieval of media streams and is derived from the *manage media* use case. The *upload media* component uploads media into the system (from an originating request in the *profiling and actions* subsystem described above). Since there are many different media formats available, the framework converts all media into a uniform format before storage in the *media streams* repository. Then, a thumbnail is created for the stream in the *thumbnails* repository. The action of uploading is stored as a user action in the user details to record whom has uploaded this file. The media streams are linked to the metadata by storing their path in the MediaOccurrence element.

The *retrieve media* component enables users to search against tags in the system and find matching media streams. The results are ranked according to frequency of occurrence via the *rank search results* process. The *rank top media* process searches the model for the top media added or updated in the system and presents them to the user so that they can be aware of current media trends within the system. Both processes utilise the *retrieve thumbnails* process to display the thumbnails inline with the results.

3.2 The MC^2 Online Service

An online service has been implemented using rapid application prototyping as a proof of concept of the framework and as a service for the community of users that participated in the experiments. The online service allows MC^2 to progress from a theoretical framework into a real functioning system and also to see it in action. MPEG-7 has been used for metadata modelling. Figure 3.5 illustrates the technologies used to develop this MC^2 service. All examples in this section are from a single video that has been deployed in the service: a 5 minute clip from the Japanese manga series, *Naruto*, where Naruto combats a strong enemy in a jungle and uses his secret move, the Shadow Replications.

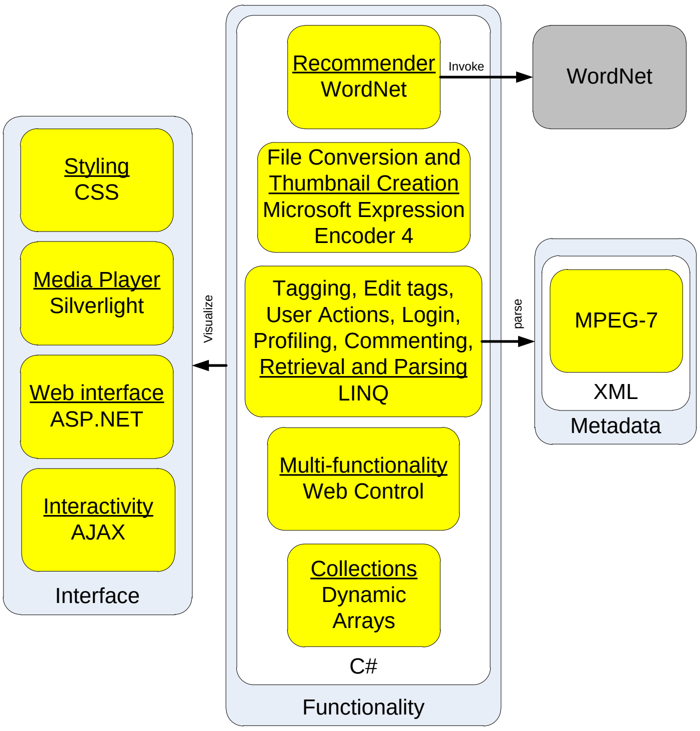


Figure 3.5: Tools and technologies used for implementation of MC²

3.2.1 MPEG-7 Profile

As stated previously, all events are modelled in MPEG-7 using the Event DS (through SemanticBase DS). Figure 3.6 presents the implementation of an event. The tag is stored under the Label and Definition, while the related media stream is indicated under the MediaOccurrence, and SubInterval represents the start time and the duration of the event. A new unique ID is created (in this case, id4) for each tag. The AgentObject DS used for people, the Object DS used for objects, the SemanticTime DS used for semantic time, the SemanticPlace DS used for locations and the SemanticState DS used for properties, all follow a similar format but with a different value for *xsi:type* and thus are not shown here to save repetition. The Object DS is also used for modelling music tags and it too follows a similar pattern, however the Creation DS and RelatedMaterial DS may also be attached to the Object DS for more detailed content modelling of music. The Creation DS includes metadata regarding the composer of the music while the RelatedMaterial DS includes more low-level audio information regarding the piece of music. Event DS is also used to model Noise tags and follows the same implementation as a regular event tag.

Figure 3.7 presents a sample conversation being represented in MC^2. As mentioned before, conversations are not specifically supported by MPEG-7, so SemanticState DS is used for this purpose. The Attribute name of the AttrubuteValuePair is set to “Conversation” and the text of the conversation is stored as the TextValue for each media occurrence. An AgentOf semantic relation connects the conversation to its speaker. When modelling conversations the Label and Property elements of the SemanticState DS are unused, which is not the case when SemanticState DS is used to model properties. As mentioned in the previous section every action in the system is stored in the user profile using the UserActionList DS. Figure 3.8 illustrates a sample where the user has created a new tag. The date and time of the action is stored along with an XPath to the actual tag. The *semantic awareness* challenge resulted in a need to track the changes made on tags. Figure 3.9 presents a sample of the Archive in the MC^2 service where the tag was first added without time points and the time points were then added at a later date. The collection keeps the original tag ID but adds a versioning index to the ID of the tags it retains. The previous figure shows how UserActionList keeps a reference to the tag ID under the Collections DS rather than the original tag so that the service can track back which version of the tag was created/updated by each user.

```
<SemanticBase id="id4" xsi:type="EventType">
   <Label>
      <Name>event-id4</Name>
   </Label>
   <Definition>
      <FreeTextAnnotation>
         flying
      </FreeTextAnnotation>
   </Definition>
   <MediaOccurrence>
      <MediaLocator>
         <MediaUri>234.wmv</MediaUri>
      </MediaLocator>
      <Mask xsi:type="TemporalMaskType">
         <SubInterval>
           <MediaTimePoint>T00:01:27</MediaTimePoint>
           <MediaDuration>PT0M02S</MediaDuration>
         </SubInterval>
      </Mask>
   </MediaOccurrence>
</SemanticBase>
```

Figure 3.6: Events in MC^2

```xml
<SemanticBase id="id256" xsi:type="SemanticStateType">
   <Label>
      <Name> </Name>
   </Label>
   <Property>
      <Name> </Name>
   </Property>
   <MediaOccurrence>
      <MediaLocator>
         <MediaUri>234.wmv</MediaUri>
      </MediaLocator>
      <Mask xsi:type="TemporalMaskType">
         <SubInterval>
            <MediaTimePoint>T00:09:42</MediaTimePoint>
            <MediaDuration>PT0M00S</MediaDuration>
         </SubInterval>
      </Mask>
   </MediaOccurrence>
   <AttributeValuePair>
      <Attribute>
         <Name>Conversation</Name>
      </Attribute>
      <TextValue>I'll show you a secret combat move…</TextValue>
   </AttributeValuePair>
</SemanticBase>
```

Figure 3.7: Conversations in MC^2.

```xml
<UserActionList>
   <ActionType href="urn:mc2:cs:UserActionCS:2008:Create"></ActionType>
   <UserAction>
      <ActionTime>
         <GeneralTime>
            <TimePoint>2011-5-4T21:57:46</TimePoint>
            <Duration>PT10S</Duration>
         </GeneralTime>
      </ActionTime>
      <ProgramIdentifier>234.wmv</ProgramIdentifier>
      <ActionDataItem href="video3.mp7.xml#id4-1"></ActionDataItem>
   </UserAction>
</UserActionList>
```

Figure 3.8: User actions in MC^2

```xml
<Collection id="id4" xsi:type="MixedCollectionType">
   <SemanticBase id="id4-1" xsi:type="EventType" xmlns="urn:mpeg:schema:2001">
      <Label>
         <Name>event-id4</Name>
      </Label>
      <Definition>
      <FreeTextAnnotation>
        flying
      </FreeTextAnnotation>
      </Definition>
      <MediaOccurrence>
         <MediaLocator>
            <MediaUri>234.wmv</MediaUri>
         </MediaLocator>
         <Mask xsi:type="TemporalMaskType" />
      </MediaOccurrence>
   </SemanticBase>
   <ContentRef href="userlist.mp7.xml#userid13"></ContentRef>
</Collection>
<Collection id="id4" xsi:type="MixedCollectionType">
   <SemanticBase id="id4-2" xsi:type="EventType" xmlns="urn:mpeg:schema:2001">
      <Label>
         <Name>event-id4</Name>
      </Label>
      <Definition>
         <FreeTextAnnotation>
            flying
         </FreeTextAnnotation>
      </Definition>
      <MediaOccurrence>
         <MediaLocator>
            <MediaUri>234.wmv</MediaUri>
         </MediaLocator>
         <Mask xsi:type="TemporalMaskType">
            <SubInterval>
              <MediaTimePoint>T00:01:27</MediaTimePoint>
              <MediaDuration>PT0M02S</MediaDuration>
            </SubInterval>
         </Mask>
      </MediaOccurrence>
   </SemanticBase>
   <ContentRef href="userlist.mp7.xml#userid13"></ContentRef>
</Collection>
```

Figure 3.9: Archiving and Semantic Awareness in MC^2

3.2.2 Functionality

The functionality of the system was developed in Microsoft C# due to the powerful and extensive libraries and Web compatibility. The recommender works in conjunction with WordNet for database access, which is a large lexical database of English developed by Princeton University (Princeton University, 2010). Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts is widely used for spell checking and phrase recommendation by researchers.

The file conversion and thumbnail creation process are implemented using Microsoft Expression Encoder 4 (Microsoft, 2012a) which will encode any audio and video into WMV format which is the acceptable format for media playback in Silverlight (discussed further below). During the conversion process, Expression Encoder will also create a thumbnail as long as it is set in the *jobfile* that has been created for the process. Another strong capability of Expression Encoder is that it can convert audio files into videos with blank or pre-set visual elements, therefore enabling users to utilise the same Web-player for playing videos and audios. In order to use Expression Encoder 4, the encoder class will invoke "Encoder.exe" as a new hidden process and adds a Jobfile as input argument for it. Jobfile is an XML file that can be customised with specifications of the conversion process. Figure 3.10 presents the sample Jobfile template.

```
<?xml version="1.0" encoding="utf-16"?>
<JobFile Version="1.0">
  <Job HtmlTemplate="App:\Templates\en\Pilot:Pilot"
      OutputDirectory="#outputdir#"
      SaveJobFile="True"
      AppendJobID="False"
           Log="On">
    <MediaFiles>
      <MediaFile  FileType="Audio, Video"
                  ThumbnailMode="FirstFrame"
                  ThumbnailSize="160, 120"
                  MarkerThumbnailSize="36, 28">
        <Markers>
        </Markers>
      </MediaFile>
    </MediaFiles>
  </Job>
</JobFile>
```

Figure 3.10: An XML Jobfile template

The main functionality of the MC^2 service is the tagging capabilities. As mentioned before all the information of this system is stored in MPEG-7 which is an XML based tool. Therefore, any information storage or retrieval action of the service relies on the XML-based metadata. The parsing algorithm implemented in this system for storing and retrieving data was developed using the

Microsoft LINQ (Language-Integrated Query) library (MSDN, 2010). It offers developers a way to query data using strongly-typed queries and strongly-typed results, therefore making the parsing operation faster and more efficient. There are three classes responsible for metadata parsing, creating and editing: MP7Manage, UserAccount and ArchiverCollections. The MP7Manage class deals with adding, retrieving and editing tags using LINQ. This class creates a new XElement that is a clone of the template tag, setting its attributes and adding/replacing it in the MPEG-7 content model. Figure 3.11 presents sample code for adding events. The UserAccount class is responsible for adding actions. The action method of this class is called by the first class at the end of add or edit methods and Figure 3.12 presents sample code of an action being added. The ArchiverCollections class caters for dealing with archiving and is called at the end of nearly every method of the first class to store the changes (Figure 3.13). As the MC^2 service generates and operates on large collections, the search results, tag previews, and user area results all need to be stored in collections. The dynamic arrays provided by C# create a suitable medium for storing and operating on large collections of dynamic sizes.

```
public string AddNewEvent(string name, string fn, string tp, string dur)
    {if (!isTaken(name, "EventType"))
           {string idd = GetID();
            var query = from c in MP7.Descendants
               ("{urn:mpeg:mpeg7:schema:2001}SemanticBase")
            where (string)c.Attribute
              ("{http://www.w3.org/2001/XMLSchema-instance}type") == "EventType"
            select c;
            XElement temp = new XElement(query.First());
            temp.Descendants("{urn:mpeg:mpeg7:schema:2001}Name")
              .First().Value = name;
            temp.Descendants("{urn:mpeg:mpeg7:schema:2001}MediaUri")
              .First().Value = fn;
            temp.Descendants("{urn:mpeg:mpeg7:schema:2001}MediaTimePoint")
              .First().Value = PrepareTime(tp);
            temp.Descendants("{urn:mpeg:mpeg7:schema:2001}MediaDuration")
              .First().Value = PrepareDuration(dur);
            temp.Attribute("id").Value = idd;
            query.First().AddAfterSelf(temp);
            MP7.Save(XMLFileName);
            arcc.CreateArchive(temp);
            return idd;}
        else {return "-1";}}
```

Figure 3.11: Adding an event in C# using LINQ

```
public static void AddAction
    (string spath, string acttype,string userid,string fn, string elemid)
       {ServerPath = spath;
        string XmlFileName = ServerPath + GetUserFilePath(userid);
        XElement MP7 = XElement.Load(XmlFileName);
        var query = from c in MP7.Descendants
           ("{urn:mpeg:mpeg7:schema:2001}UserActionList")
        select c;
        XElement temp = new XElement(query.First());
        temp.Descendants("{urn:mpeg:mpeg7:schema:2001}ActionType")
              .First().Attribute("href").Value =
              "urn:mc2:cs:UserActionCS:2008:" + acttype;
        temp.Descendants
              ("{urn:mpeg:mpeg7:schema:2001}TimePoint").First().Value =
            DateTime.Now.Year + "-" + DateTime.Now.Month +
            "-" + DateTime.Now.Day + "T" +
            DateTime.Now.ToLongTimeString();
        temp.Descendants
              ("{urn:mpeg:mpeg7:schema:2001}ProgramIdentifier")
              .First().Value = fn;
        temp.Descendants
              ("{urn:mpeg:mpeg7:schema:2001}ActionDataItem")
            .First().Attribute("href").Value =
            fn.Replace(".wmv", ".mp7.xml#") + elemid;
        query.First().Parent.FirstNode.AddAfterSelf(temp);
        MP7.Save(XmlFileName);
        if ((acttype != "Create") &&
              (!IsCreator(ServerPath, userid, elemid, fn)))
           {string crtr = FindCreator(ServerPath, elemid, fn);
            if (crtr != "")
            {SetNotification(ServerPath, fn, elemid,
                     crtr, GetUserName(userid, ServerPath), acttype);}}}
```

Figure 3.12: Adding a user action in C# using LINQ

```
public void AddChange(XElement xel0)
       {XElement xel = new XElement(xel0);
        var query = from c in Archi.Descendants()
           where ((c.Name.LocalName == "Collection") &&
               (c.Attribute("id").Value == xel.Attribute("id").Value))
        select c;
        XElement temp = new XElement(query.First());
        XElement temp2 = query.First();
        temp.Attribute("id").Value = xel.Attribute("id").Value;
        temp.Descendants("{urn:mpeg:schema:2001}ContentRef")
             .First().Attribute("href").Value = RefUserID;
        int id = int.Parse(query.Last().Descendants().
             First().Attribute("id").Value.
             Replace(xel.Attribute("id").Value + "-","")) + 1;
        xel.Attribute("id").Value = xel.Attribute("id").Value +
              "-" + id.ToString();
        temp.Descendants().First().ReplaceWith(xel);
        temp2.Parent.LastNode.AddAfterSelf(temp);
        Archi.Save(XMLFileName);
        UserAccount.AddAction(
              XMLFileName.Remove(
              XMLFileName.IndexOf("Archive\\")), "Update",
              RefUserID.Replace("userlist.mp7.xml#", ""),
             XMLFileName.Remove(0,XMLFileName.LastIndexOf("\\")+1)
              .Replace(".mp7.xml", ".wmv"), xel.Attribute("id").Value);}
```

Figure 3.13: Storing changes in the model in C# using LINQ

3.2.3 Interface

The Web interface of the service was developed in ASP.NET and stylised using CSS to ensure maximum platform compatibility, functionality and presentation. ASP.NET is a Web developing tool that is compatible with most browsers and is widely used within the Web development community.

Figure 3.14 presents the Add Tag page in the system. As can be seen the page is divided into two areas: the media player and the tagging section. To ensure the interactivity of the interface and facilitate its use while making the pages ‘light’ enough to be used easily by users, AJAX has been used where necessary to ensure a sound presentation of the interface and ease of use. In order to ensure that add/edit actions would not affect the play back of media, the player and the add/edit panels are in two separate Update Panels. In the tagging section, users choose the type of tag they want to add from a list, either fully committing themselves to adding it with all the options available, or via the quick add option to quickly add objects, events or persons. Figure 3.15 shows an example of the Edit tag page's edit panel.

To enable the Webpages to display comprehensive information and support comprehensive functionality without unnecessary redirections, and also to ensure the media stream is not stopped during user interactivity, Web controls have been used to portray different stages and options of tagging and tag editing. For each process within the add/edit components of the MC^2 framework, a separate Web control is created. This enables quick interactive responses and ensures that each type of add/edit is performed independently. Figure 3.16 shows the available tagging options in the system. Selecting a tag in the list will dynamically unload the selection Web control and load a new Web control that is responsible for adding the selected tag instead.

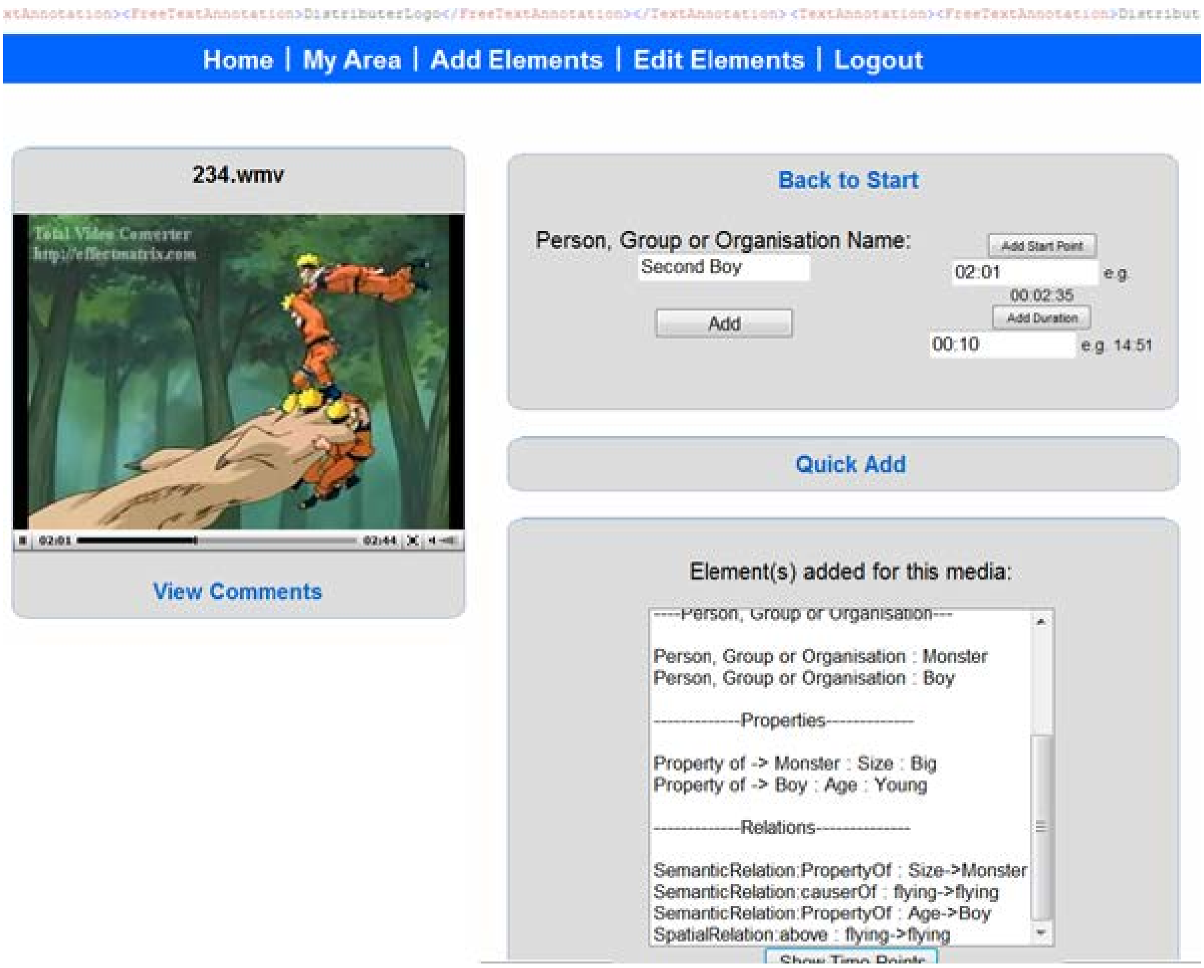


Figure 3.14: Add Tag page

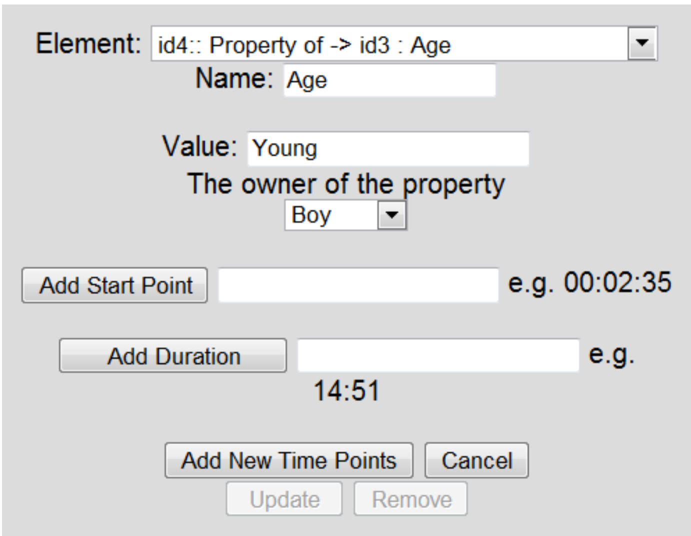


Figure 3.15: Editing a property in edit page

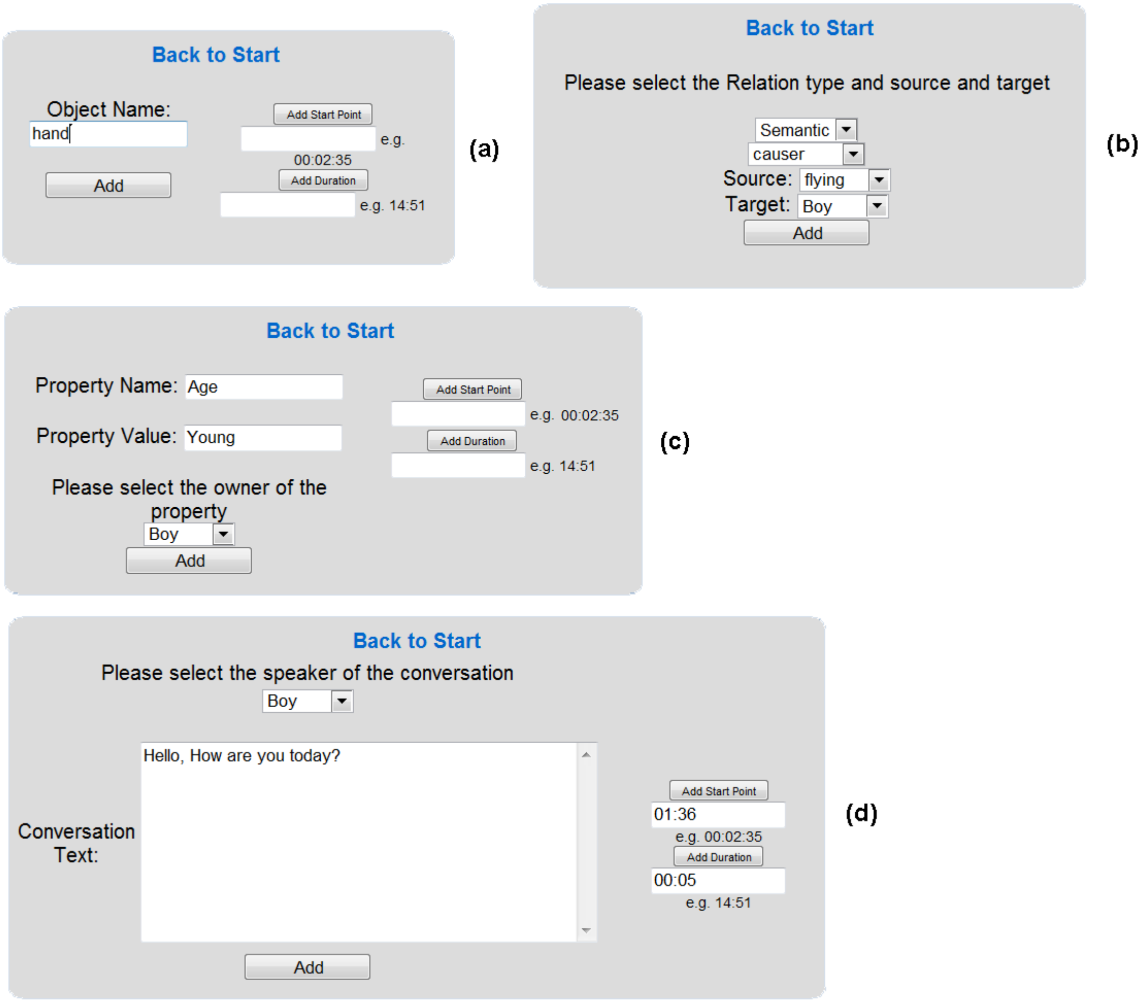


Figure 3.16: Main tags in the system:
(a) Objects, Events, People and Location (b) Relations (c) Properties (d) Conversations

The Web-based player of the service was developed using Silverlight (Microsoft, 2012b). With its XML-based format and non-complex functionality, it provides a light Web-player suitable for online streaming of media that is compatible with all major browsers. To interact with the player and extract information such as media time from the player, JavaScript functions access the player and extract the required information.

On the home page, the users have the option to upload new media or search for existing media using keywords which will present them with ranked results based on the presence of these keywords within the content model. Also, the users are presented with the top 3 most recently updated media (recently uploaded or recently modified). As mentioned previously, both search results and most recent media are found based on their models stored in MPEG-7. The search class will parse the MPEG-7 of the user actions using LINQ and will select the most recent actions and media related to them. As LINQ allows querying of XML files, the system will look for distinct new media. The search class also looks for keywords in the system and will rank each result in the set based on the times the keyword has been mentioned in the model or the number of relationships to that keyword.

3.3 Summary

The chapter contributes, first, a framework, MC^2, for modelling content collaboratively and, secondly, an online service implemented both as proof of concept of the framework and as a service to the users who participated in the experiments. The framework is based on the fuzzy cognitive model of user behaviour and was developed with use cases and grounded theory. The online service has been implemented using rapid application prototyping and all metadata has been modelled using MPEG-7. The next chapter validates MC^2's proof of concept through the same user experiment that led to the conceptual model of user behaviour.

Chapter 4: MC^2 Framework Walkthrough and Validation

MC^2 is designed and developed to strike a balance between the advantages of folksonomy and MPEG-7 tools, and also to overcome their shortcomings. In order to validate the success claims, first, MC^2 is evaluated as a tool through the same user experiment that led to the conceptual model of user behaviour and with the same groups of users and then compare the results to those of the initial experiment and, second, the MC^2 functionality is evaluated against that of folksonomy and MPEG-7 tools. The chapter starts with a walkthrough of the MC^2 online service presented as a typical "day in the life" of the service.

4.1 A day in the life of MC^2

A walkthrough of the MC^2 online service provides a comprehensive demonstration of functionality from the viewpoints of both a single user and collaborating users. Users ikBrunel, johnSmith and janeDoe have been selected for this walkthrough, these are personas adopted for demonstrating the system functionality. Figures 4.1 to 4.39 cover the walkthrough of the online service. All screenshots deploy a single 5 minute video clip from the Japanese manga series, *Naruto*. In this video clip, Naruto wakes up from sleep and combats a strong enemy in the jungle. The video was chosen because it is an eventful, generic video and in the public domain.

4.1.1 Register and Login

Interaction begins with users either logging in as in Figure 4.1 or signing up as in Figure 4.2. With the latter, once the registration information is validated, the Profiling and Actions subsystem creates an MPEG-7 profile using UserDescription DS and the user is then added to the list of users. In the example of Figure 4.2, once the user ikBrunel registers an MPEG-7 user file, ikBrunel.mp7.xml, is created and the user is then added to the list of users. ikBrunel is then directed to the login page.

Figure 4.1: Login page

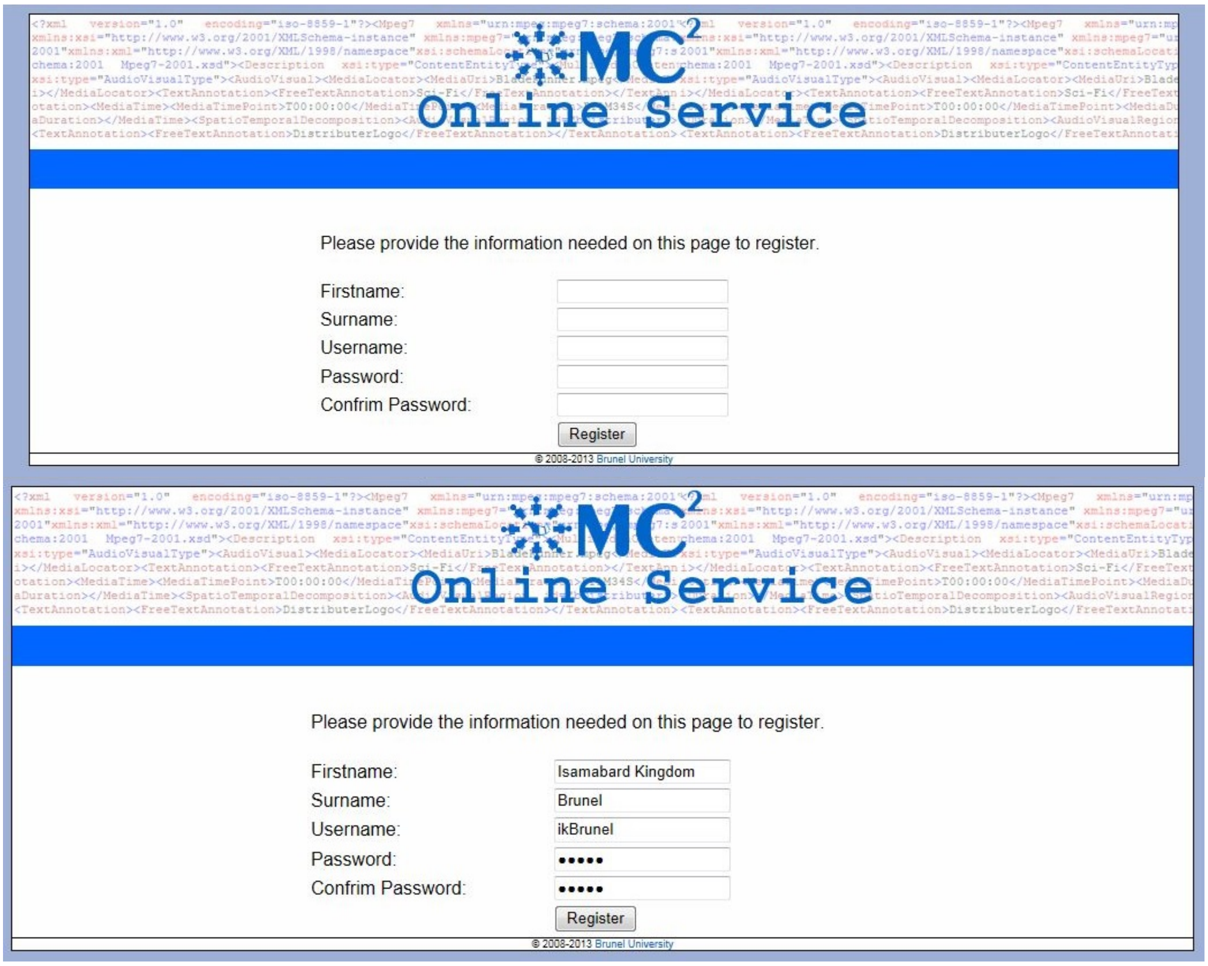


Figure 4.2: Registration page

4.1.2 Service Home Page: Search, Upload and Convert

Once the user successfully logs in, they are directed to the service home page shown in Figure 4.3 where they can either upload new media, or search existing media or choose one of the top 3 most recently updated media. The home page and all functionality related to this page is managed by the

Media Management subsystem. Figure 4.4 shows ikBrunel uploading a video. With unsupported media formats, the service converts, with the user's permission to a supported format. Figure 4.5 and Figure 4.6 illustrate this. A thumbnail of the media is also generated.

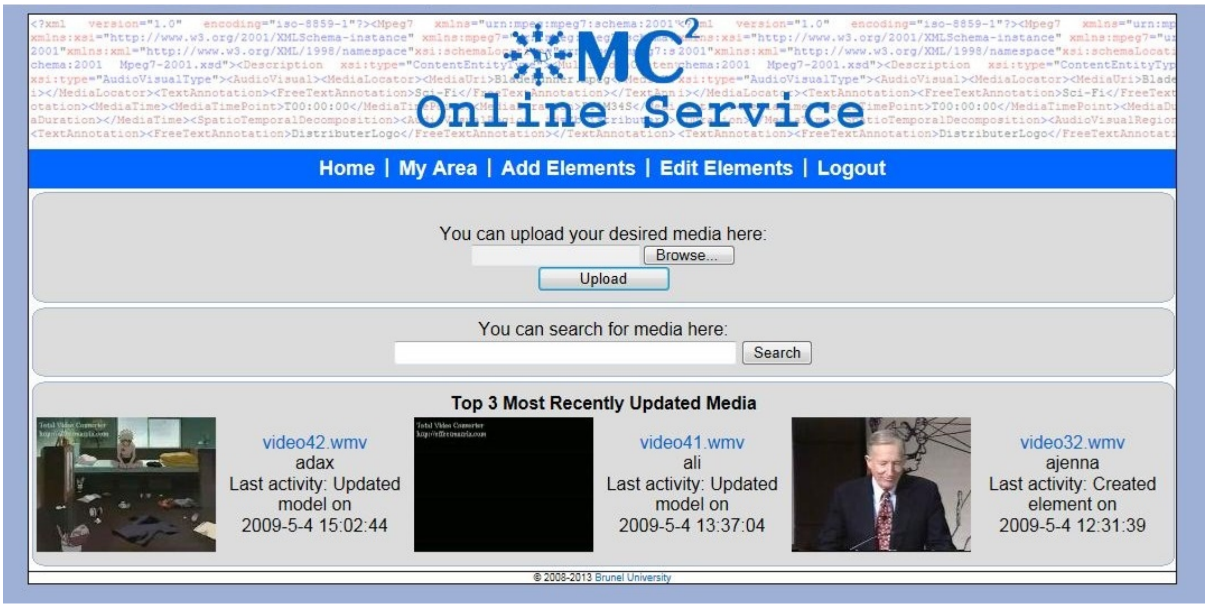


Figure 4.3: Home page

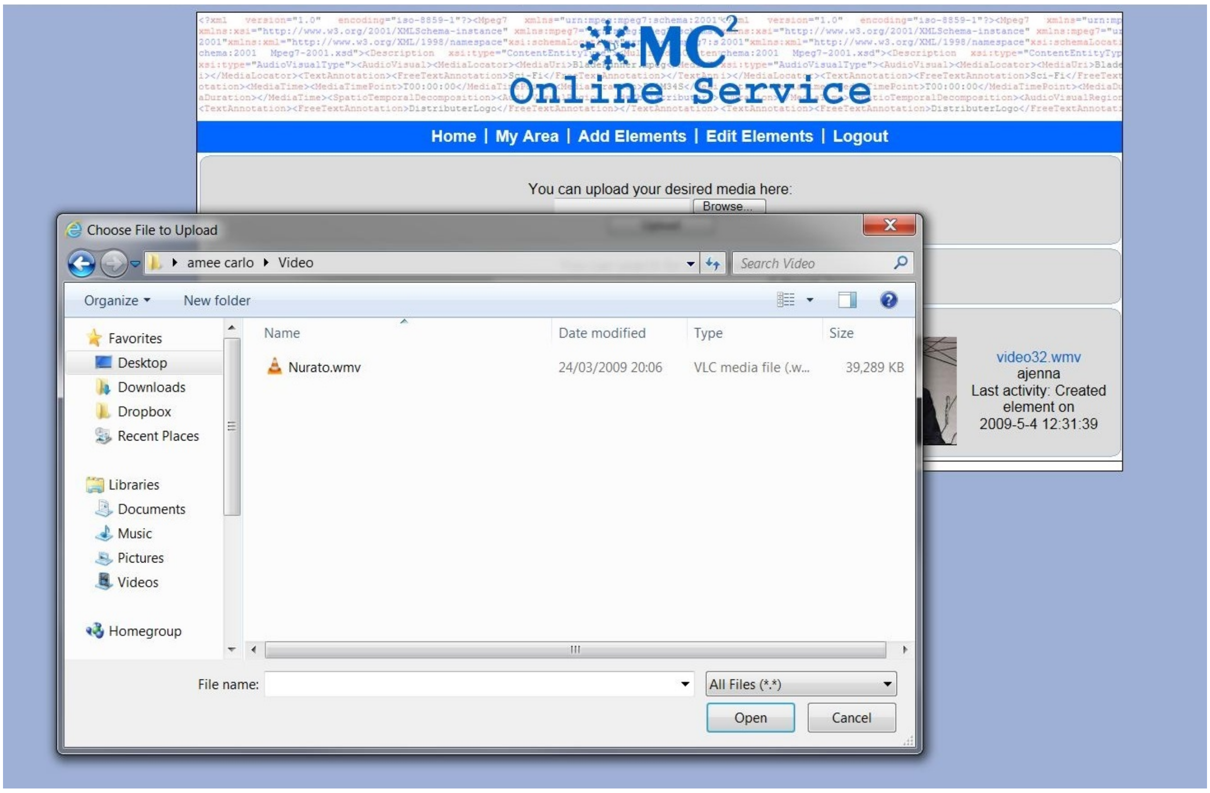


Figure 4.4: Uploading a media file

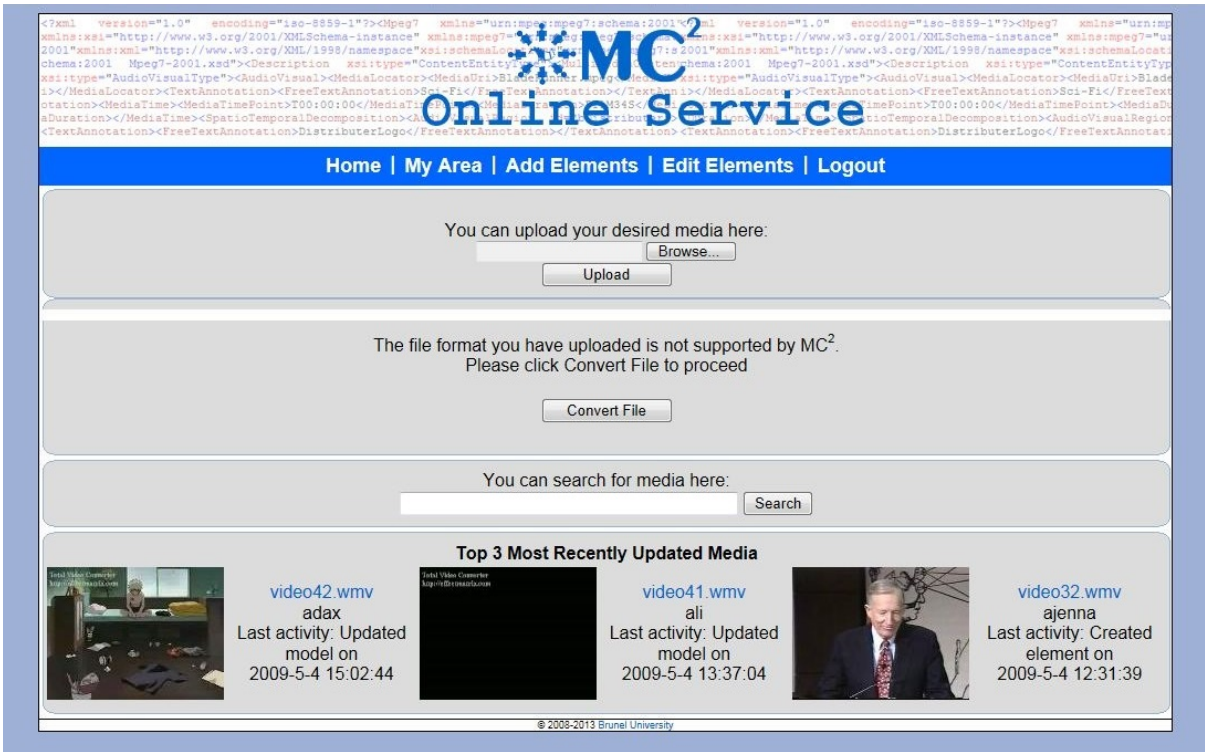


Figure 4.5: Media format conversion

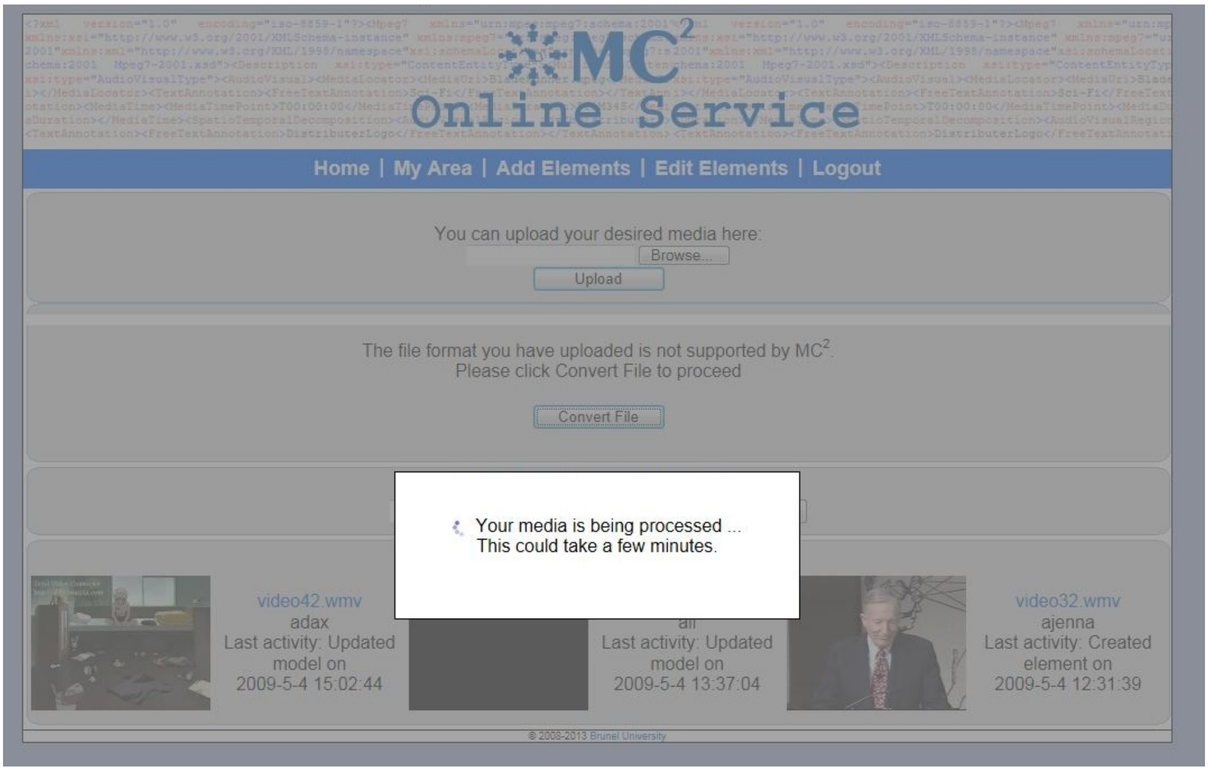


Figure 4.6: Media format conversion in progress

Uploaded media with the correct or corrected format are added to the media repository. A blank content model is created for each piece of media to store tagging information and associated comments. Changes to media are recorded to an MPEG-7 archive and declared with a new version. The service also records media uploading with a new user action, “Uploaded”, in the user model.

4.1.3 Add Element Page: Full Add

Following successful media upload, the user is directed to the “Add Element” page shown in Figure 4.7. The Viewing subsystem that partly manages this page provides a media player, whereas the Tagging subsystem provides a “Full Add” component with a list of elements users can tag and a “Quick Add” component which allows users to add objects, events or people. Tags which have been added for each piece of media are displayed. Figure 4.7 shows a page with no elements tagged yet. In contrast, Figure 4.8 shows that tags are available: Object, Event, Person, Group or Organisation, Location, Property, Conversation, Time, Relation.

Figure 4.7: Add Elements page

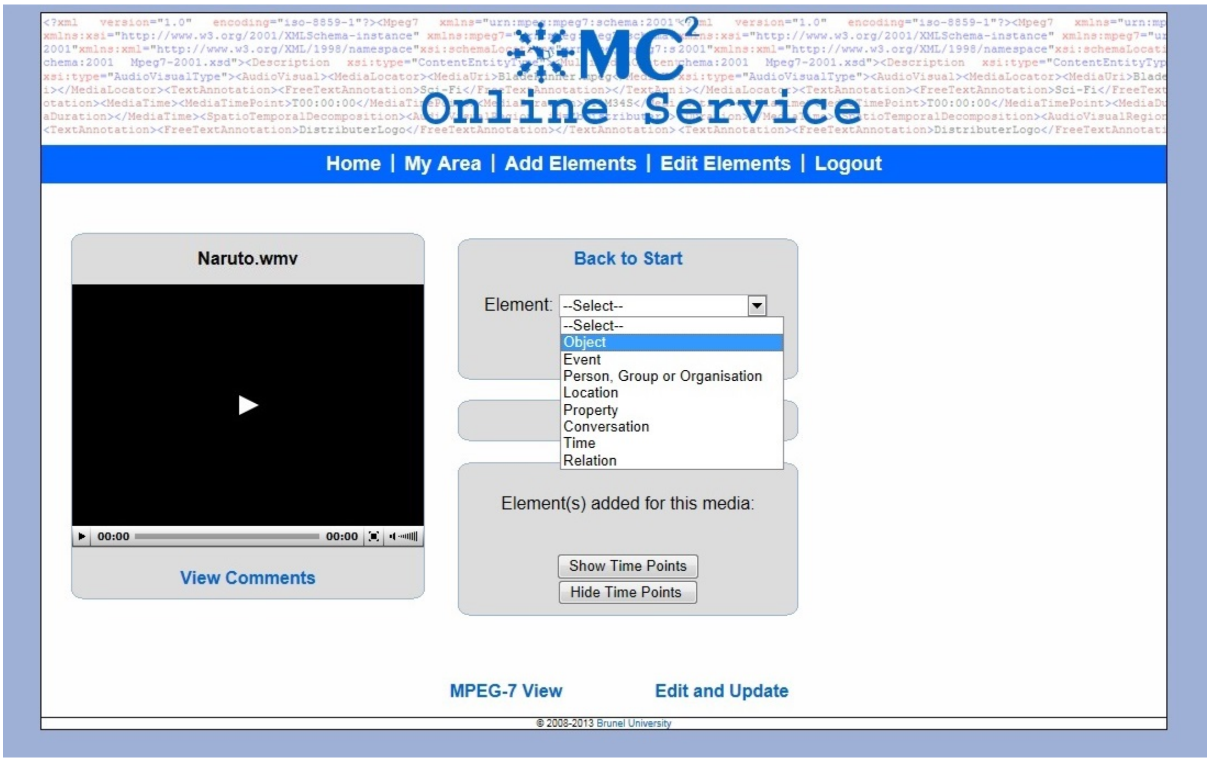


Figure 4.8: Adding Elements in progress

4.1.4 Add Object, Time Points and Recommendation

In Figure 4.9 the user tags “Hat”. In turn, the tagging subsystem adds this tag to the content model, and a record of the tag in the archives. It also adds a new user action, “Created”, in the user model. Once the object is added, the user is prompted to add Time Points for this element. This is illustrated in Figure 4.10. If the user decides to add time points they will be presented with the time points menu. Figure 4.11 shows how they can either capture the start point and duration directly from the player through the Timer component or enter them manually.

If the user chooses to add time points, these are presented with three options: add more time points, add a property or to finish with this tag. This is depicted in Figure 4.12. Once finished, the object is added to the model. This is displayed in the Elements’ box under Objects along with all other elements that have been modelled. Once an element tagging finishes, the Recommender subsystem makes suggestions from WordNet based on the tag value the user has added. This is depicted in Figure 4.13. If there are no suggestions as in Figure 4.14, this may suggest misspelling.

Figure 4.9: Adding Object

Figure 4.10: Prompting to add time points

Figure 4.11: Adding time points

Figure 4.12: Further options after adding an object

Figure 4.13: Word suggestions for "Hat"

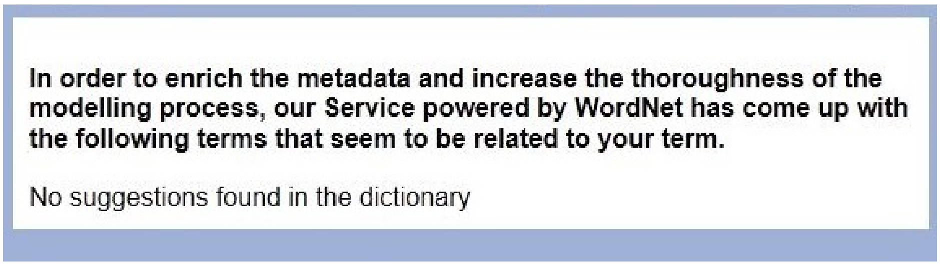


Figure 4.14: No suggestions found

Throughout the process of adding tags, free-text tagging is supported by enabling the user to add any textual string that they believe to be appropriate for describing the multimedia content. However, whereas the free-text tagging of folksonomy tools leaves such free-text tagging open to corruption, ambiguity and structure-less metadata, MC^2 provides structure on the free-text tags through the Description Schemes of the proposed MPEG-7 Profile. In this way, MC^2 yields the advantages of both free-text tagging and structured approaches while mitigating the disadvantages.

4.1.5 Add Event

The process of adding events is similar to adding an object, but this time the user decides to add a property to the event he tags. Figure 4.15 shows the user adding the event "Sitting".

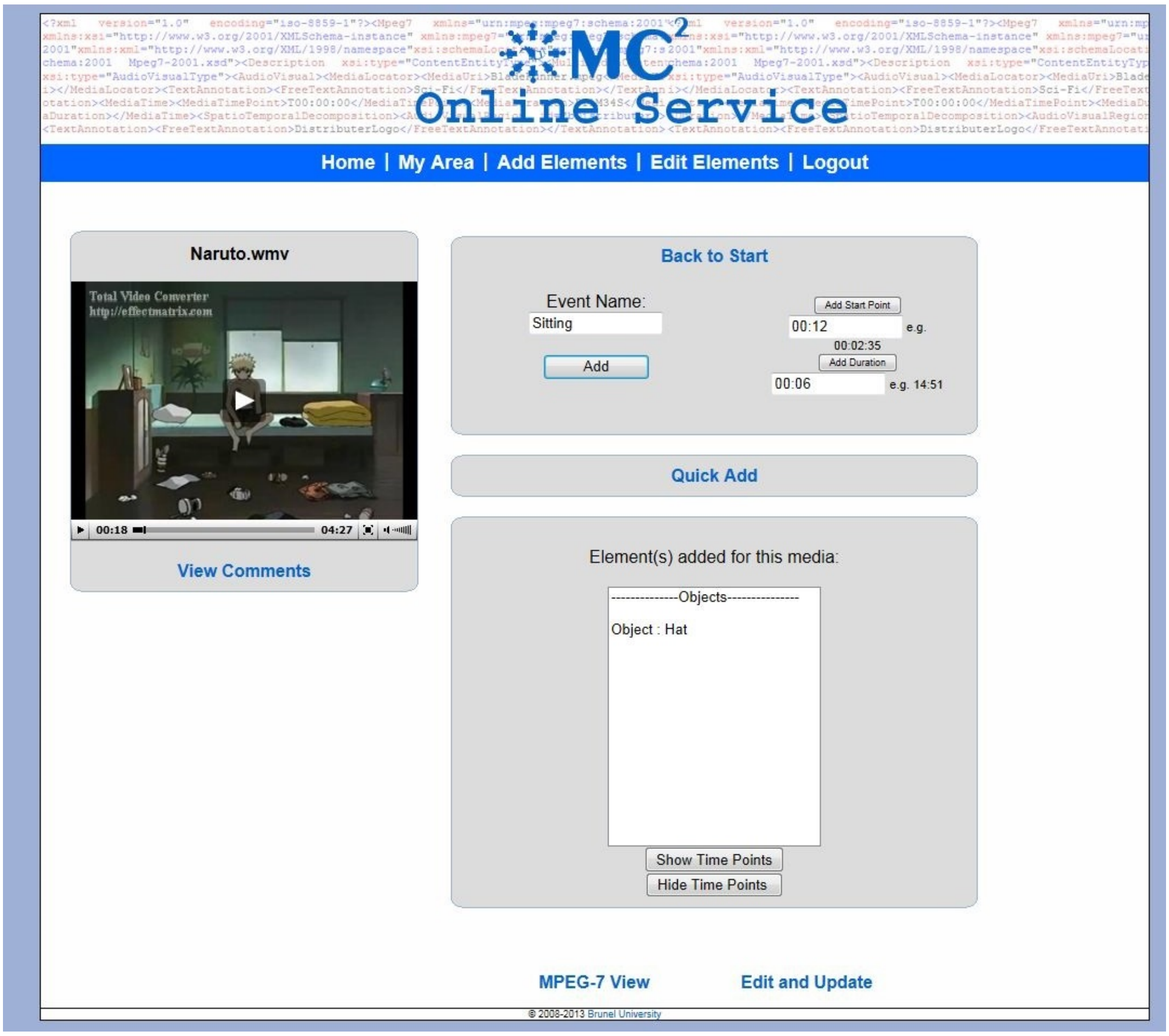


Figure 4.15: Adding Event

4.1.6 Add Property

Figure 4.16 illustrates further the add/edit property component that takes a property name and a value as inputs. In the case depicted in Figure 4.16, the user adds "state" for property name and "restless" for value. The user may change the owner of this property event if he chooses to.

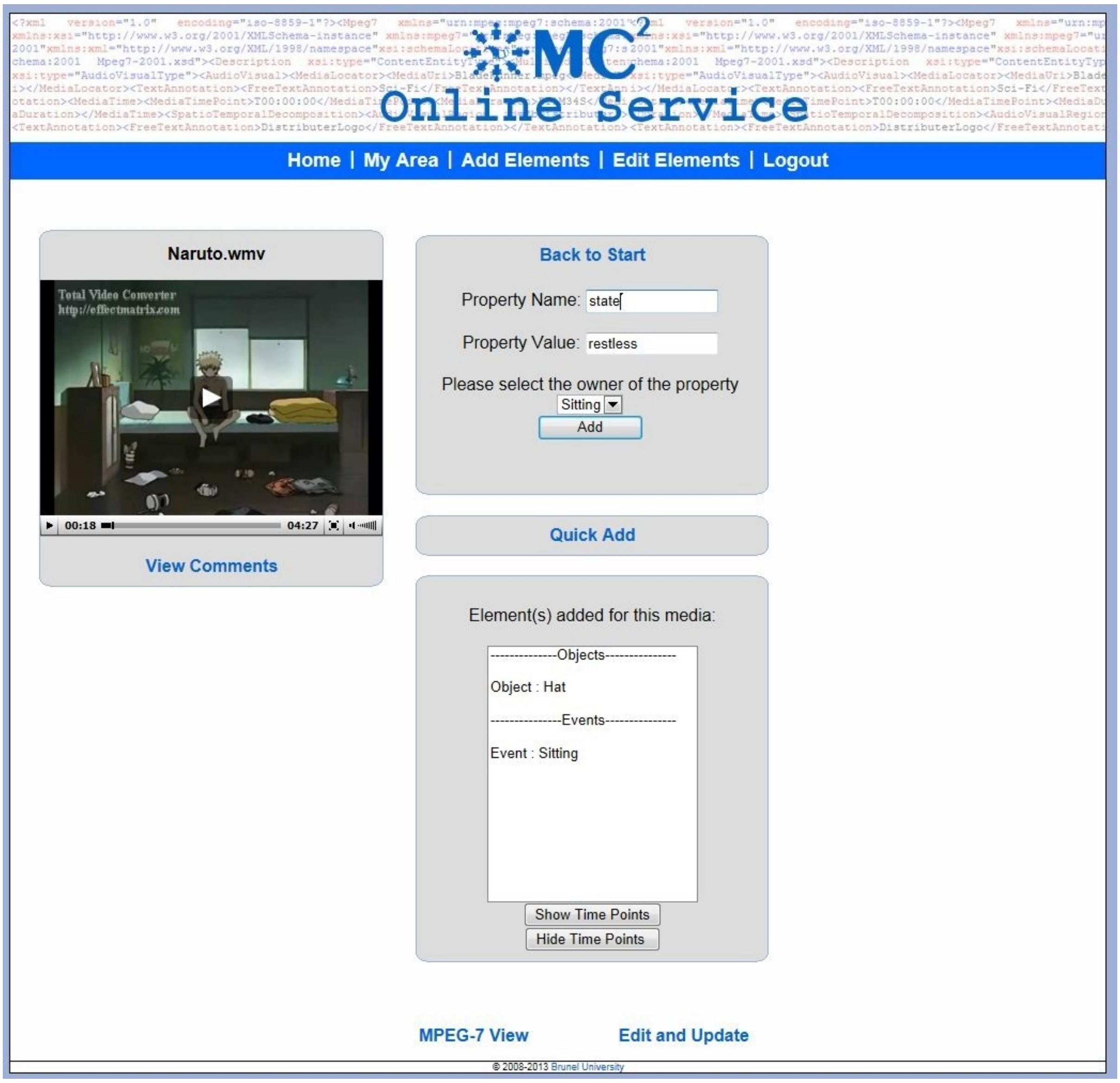


Figure 4.16: Adding property

Once a property is added, it appears in the Elements' list as shown in Figure 4.17. Furthermore, a semantic relation of the type "PropertyOf" has been added to the model. The Tagging subsystem facilitates connection between co-related tags by adding this semantic relation automatically.

4.1.7 Add Person

Tagging a Person is also a similar process to tagging Objects or Events. Figure 4.18 and Figure 4.19 illustrate how to tag a person.

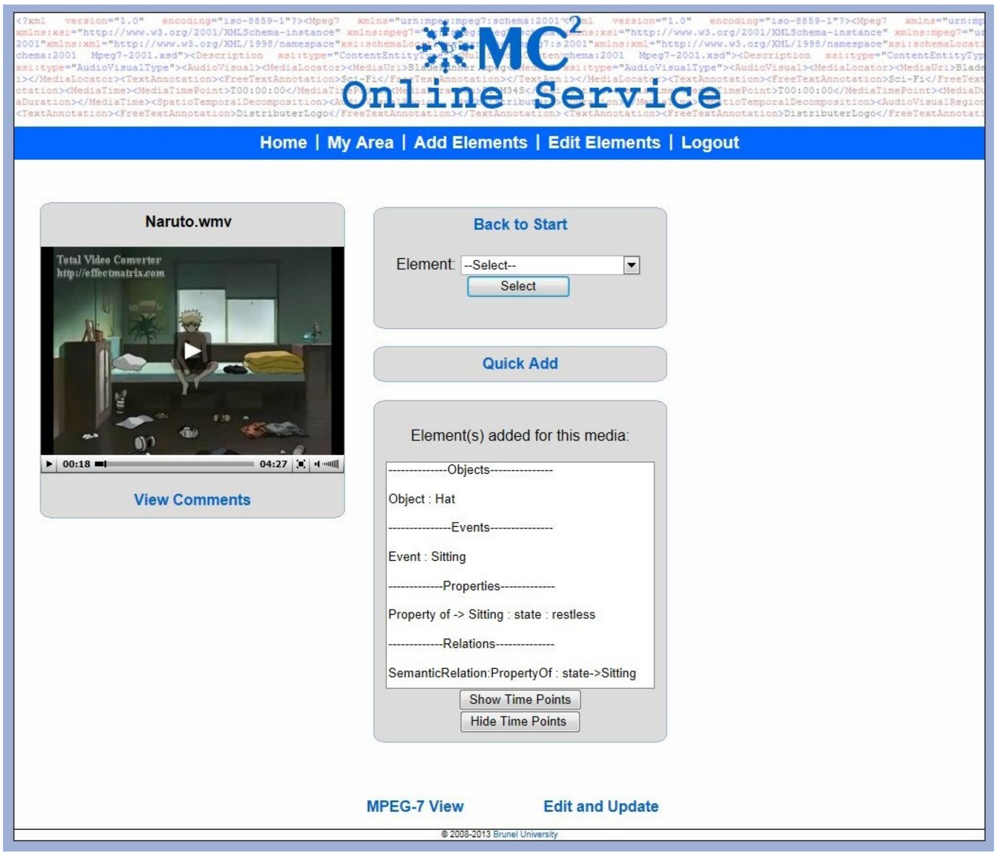


Figure 4.17: Property added

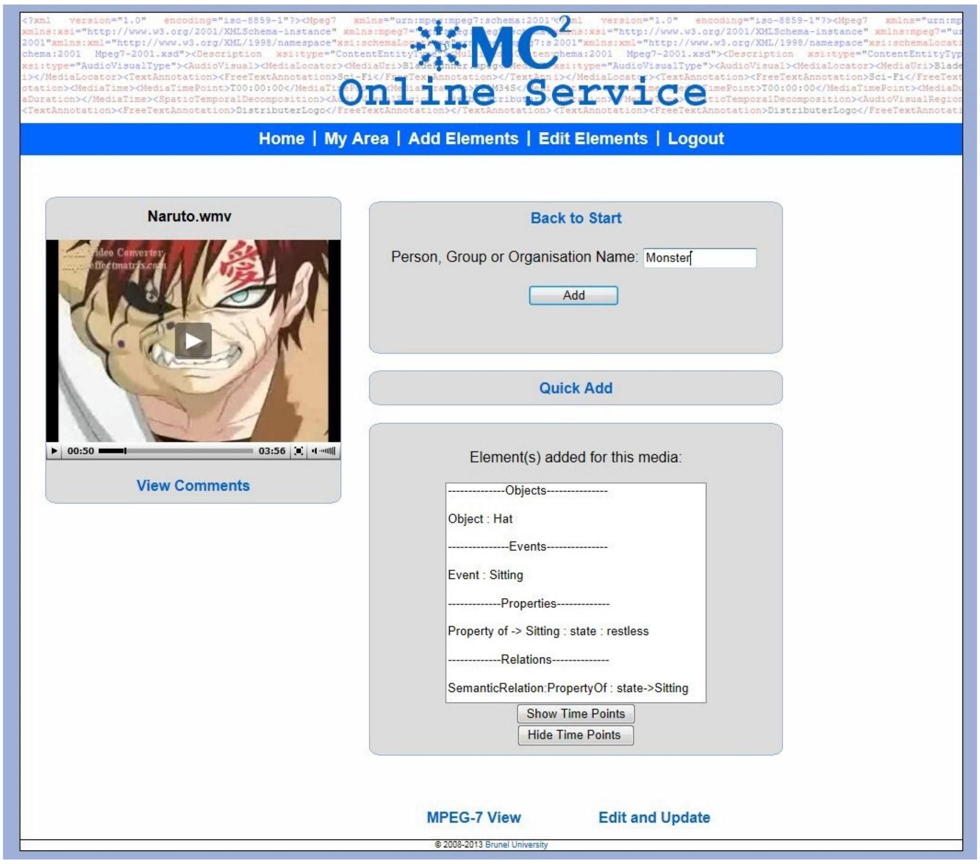


Figure 4.18: Adding a person

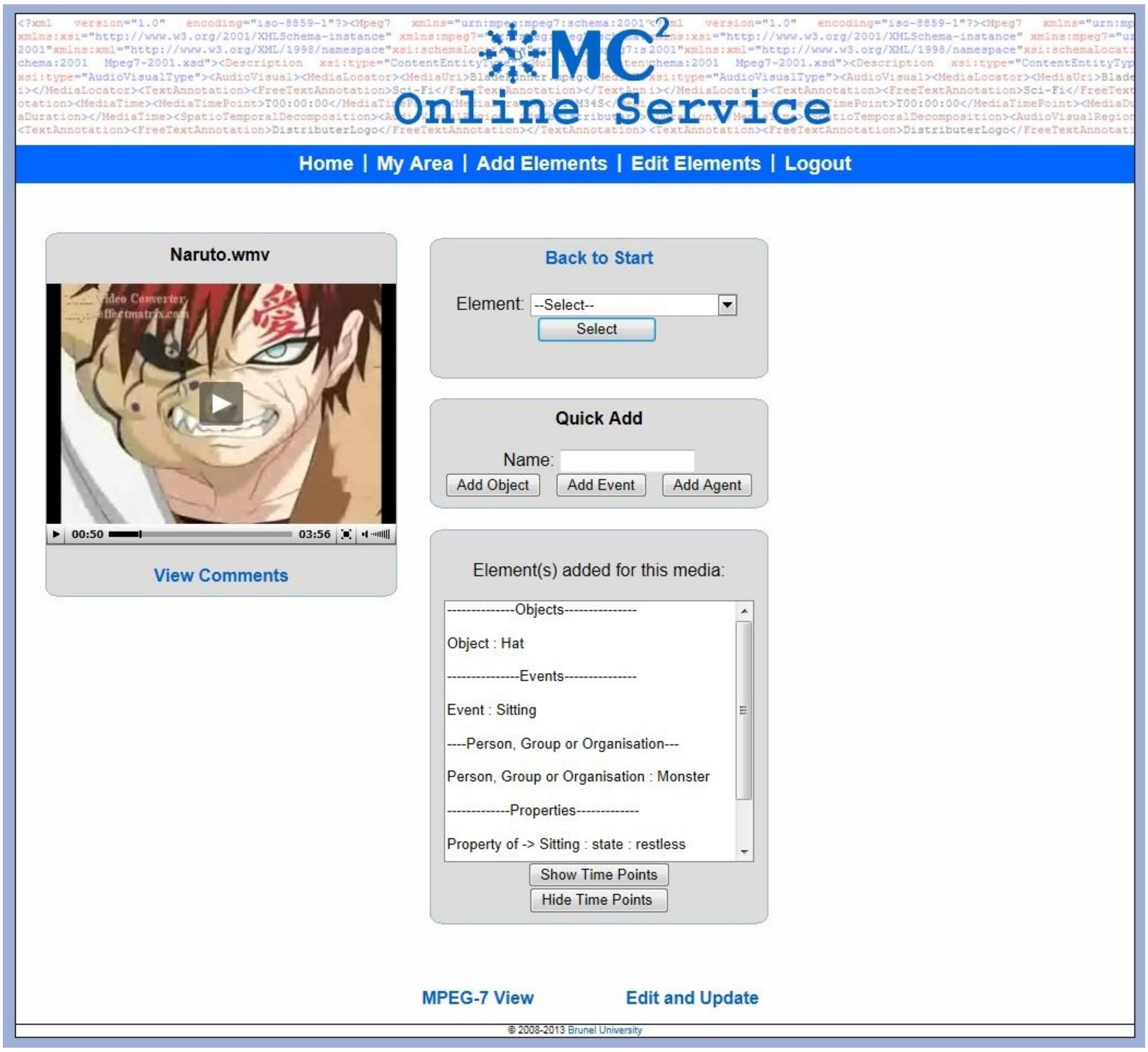


Figure 4.19: Person added

4.1.8 Show/Hide Time Points

The user can show or hide the time points for the elements at any time on the Elements' list. Figure 4.20 illustrates the case where a user chooses to show the time points for object "hat". This enables a user to check if more time points need to be added in the content model.

4.1.9 Add Conversation

The user may also add conversations. In order to achieve this, the user would need to choose a person tagged previously, and enter a conversation text. The user may also choose to add time points for this conversation. Figure 4.21 shows user ikBrunel adding a conversation.

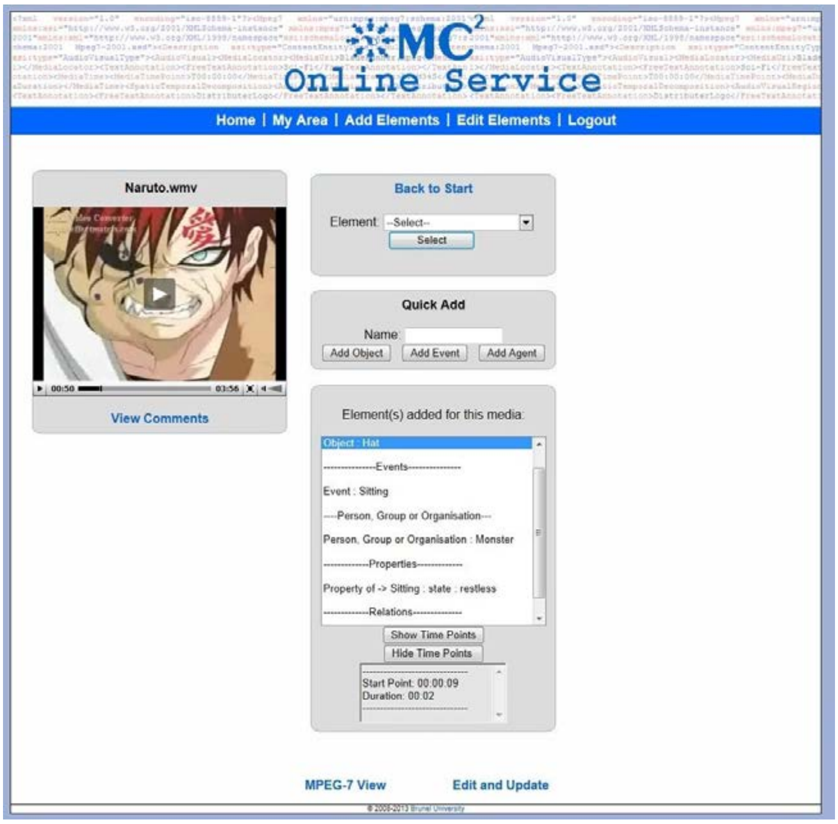


Figure 4.20: Showing time points

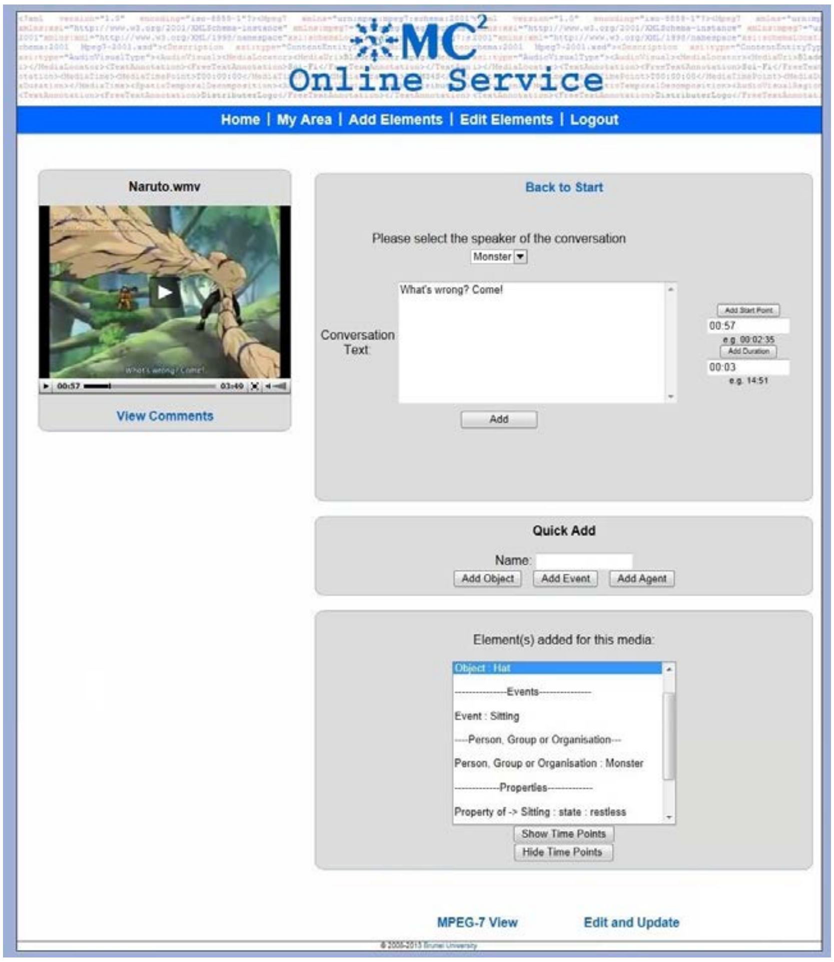


Figure 4.21: Adding conversation

The Add/Edit Conversation component records the conversation as a SemanticState DS in the content model and in the archive, and records a new user action reference to this in the user model. An added conversation is depicted on the Elements' list in Figure 4.22.

4.1.10 Add Relations: Semantic and Spatial

Users can also add relations through the Add/Edit Relation component. In Figure 4.23 a user is adding a semantic relation of the *CauserOf* type between the person "Naruto" and the event "Sitting". Once the relation is added, it can be viewed on the Elements' list. In Figure 4.24 a user is adding a spatial relation of type *above* between the person "Naruto" and the object "Bed".

Figure 4.22: Conversation added

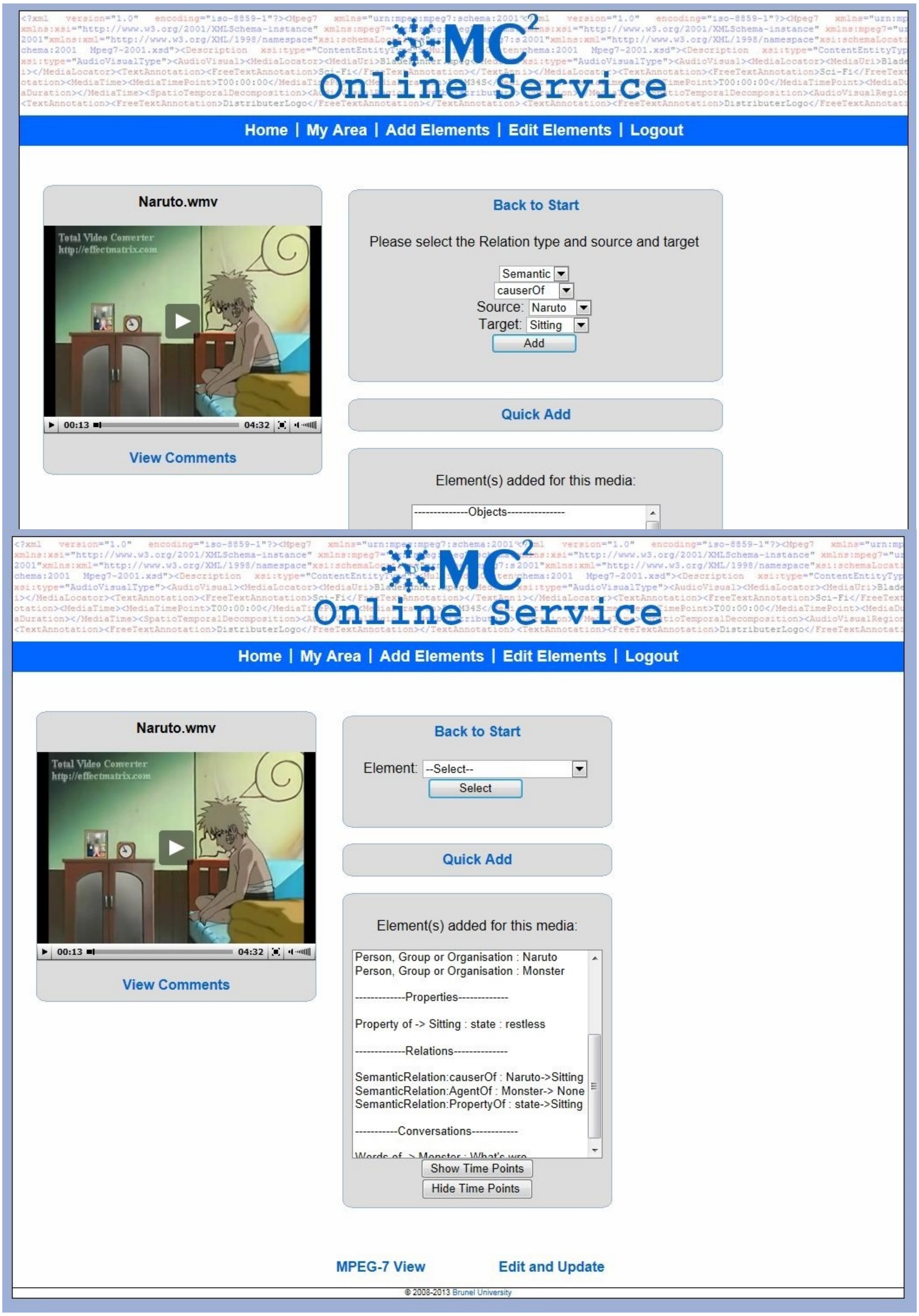


Figure 4.23: Adding a semantic relation

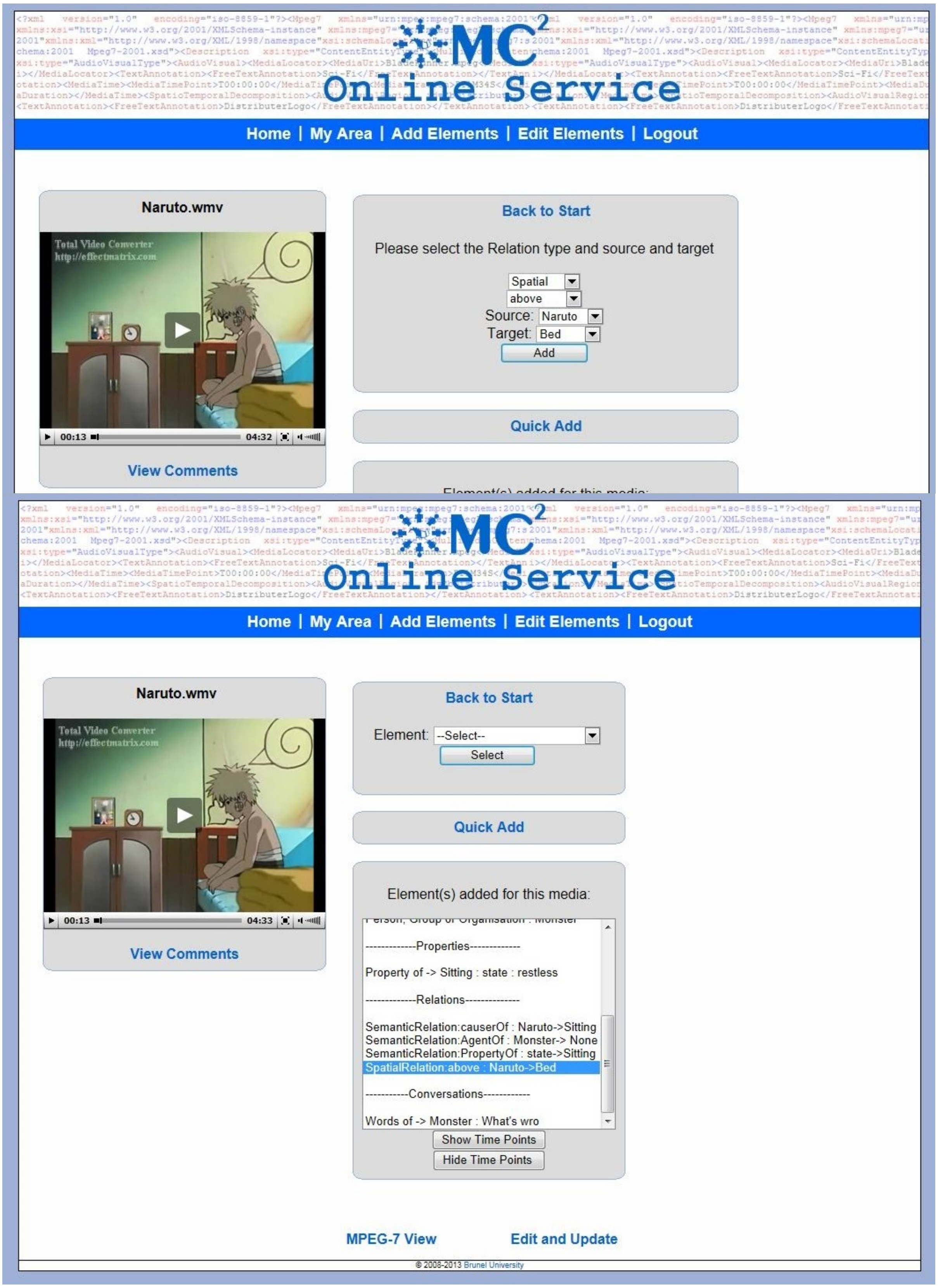


Figure 4.24: Adding a spatial relation

4.1.11 User Area: My Area

A user may view the media streams they have added or tagged by going to "My Area". Figure 4.25 shows that ikBrunel has uploaded one piece of media and has tagged two other. The page depicted in Figure 4.25 is managed by the User Area component of the Profiling and Actions subsystem. This component retrieves the relevant metadata from the user model.

4.1.12 Service Home Page: Most Recently Updated Media

The retrieve media component returns a list of the most recently updated media streams. For instance, other users will be able to see the three videos added or tagged by ikBrunel. Figure 4.26 shows johnSmith's view of home page.

4.1.13 Service Home Page: Search

Figure 4.27 depicts johnSmith searching for a video using the keyword "Bed". The retrieve media subsystem locates the elements which it then ranks based on the number of keyword occurrences in the model. The search results also include the number of occurrences of the keyword.

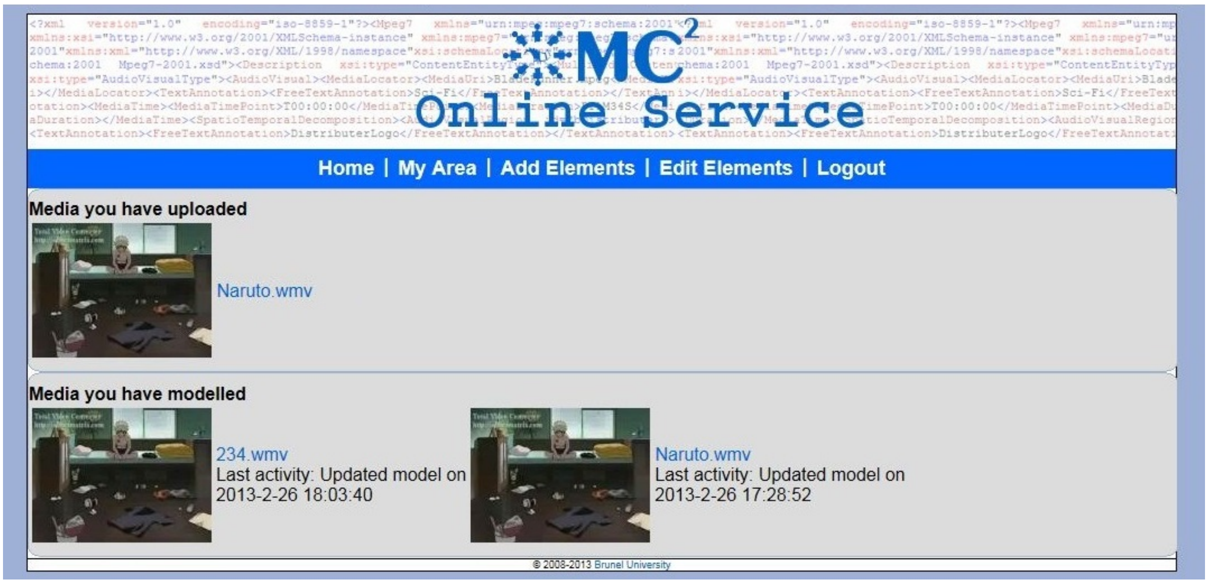


Figure 4.25: My Area

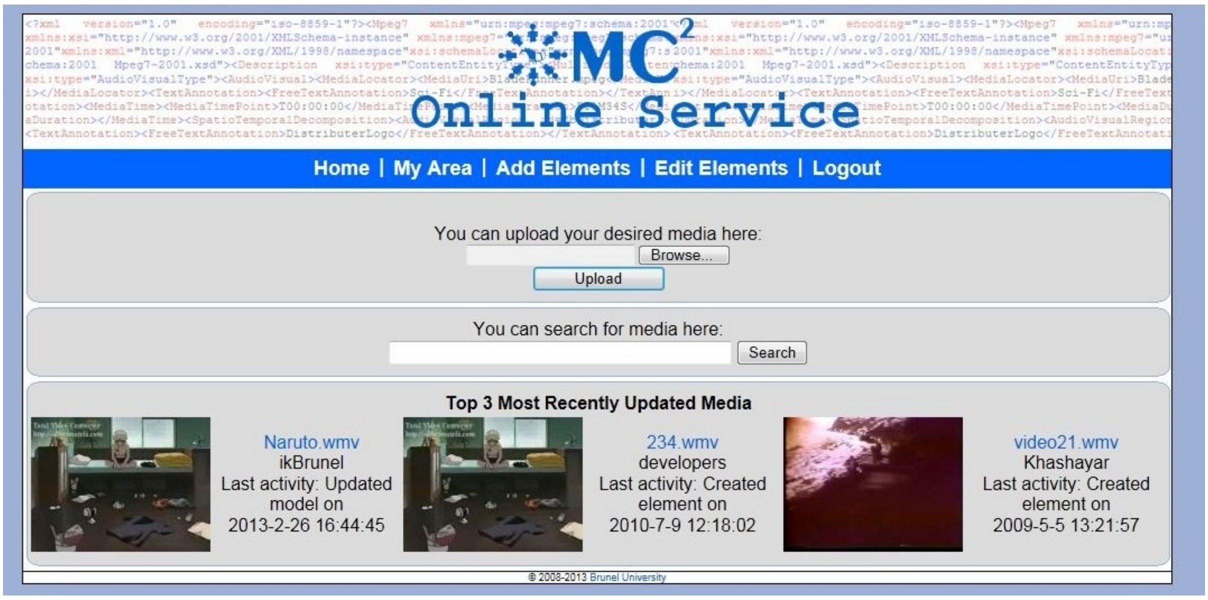


Figure 4.26: Home page for johnSmith

Figure 4.27: Search results

4.1.14 Collaboration: Tag Update by another User, Update Notification

A second user, johnSmith, decides to view “Naruto.wmv”, uploaded and tagged by ikBrunel. After reviewing they decide to add a new event, “jumping”, to the model as shown in Figure 4.28. They also decide to edit the “sitting” event tagged by ikBrunel. In order to achieve this, the user visits the Edit Elements page on which all the elements of the content model are loaded onto a drop down list, as shown on Figure 4.29, including the element id, tag type and value. In Figure 4.30, johnSmith selects “sitting” along with its time points for editing. On completion, both the content model and archive are updated and a new user action, “Updated”, is added to the user model. The versioning subsystem notices that johnSmith was not the creator of this tag and notifies ikBrunel as shown on Figure 4.31 that a tag they added has been edited. ikBrunel may choose to accept or reject this.

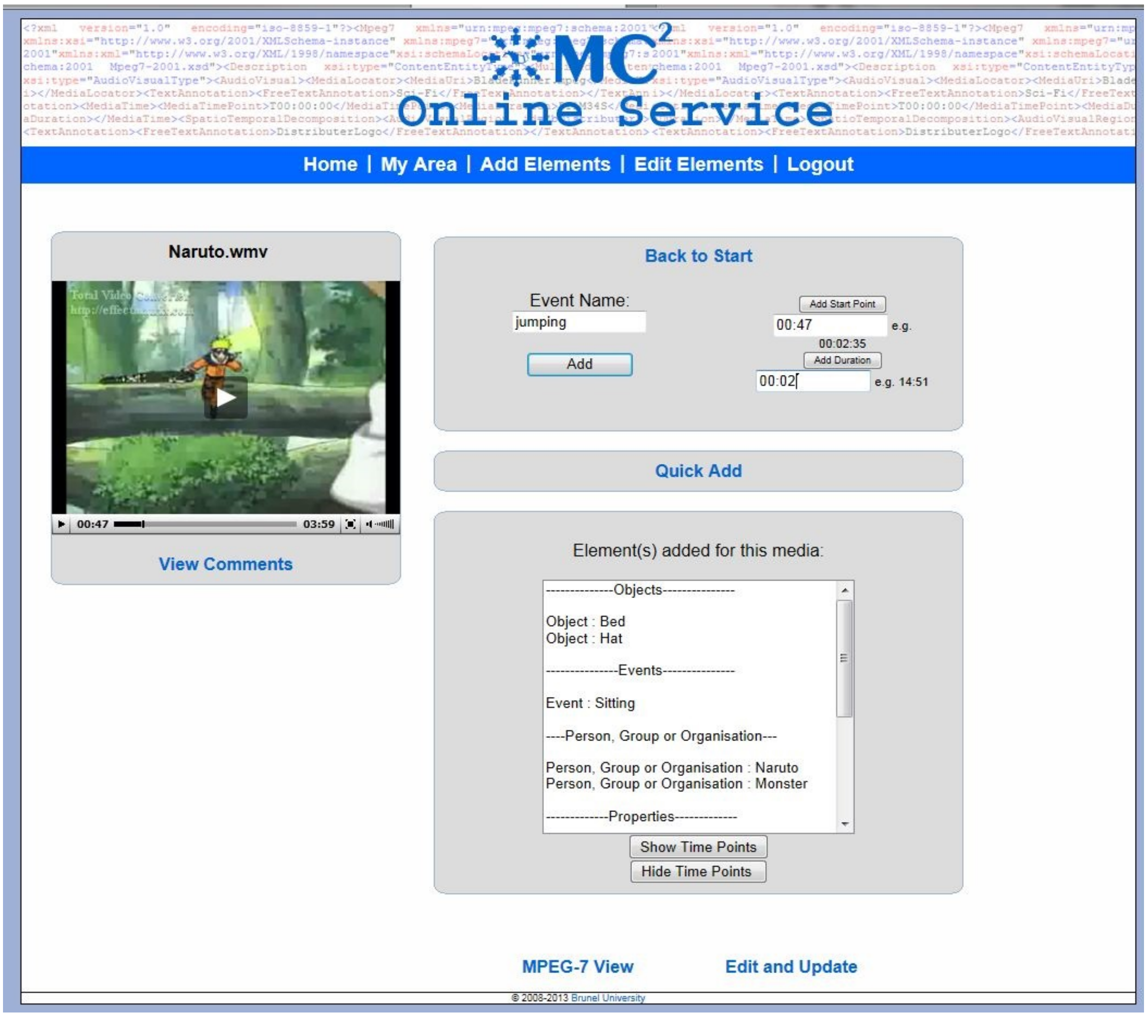


Figure 4.28: johnSmith tags a new event

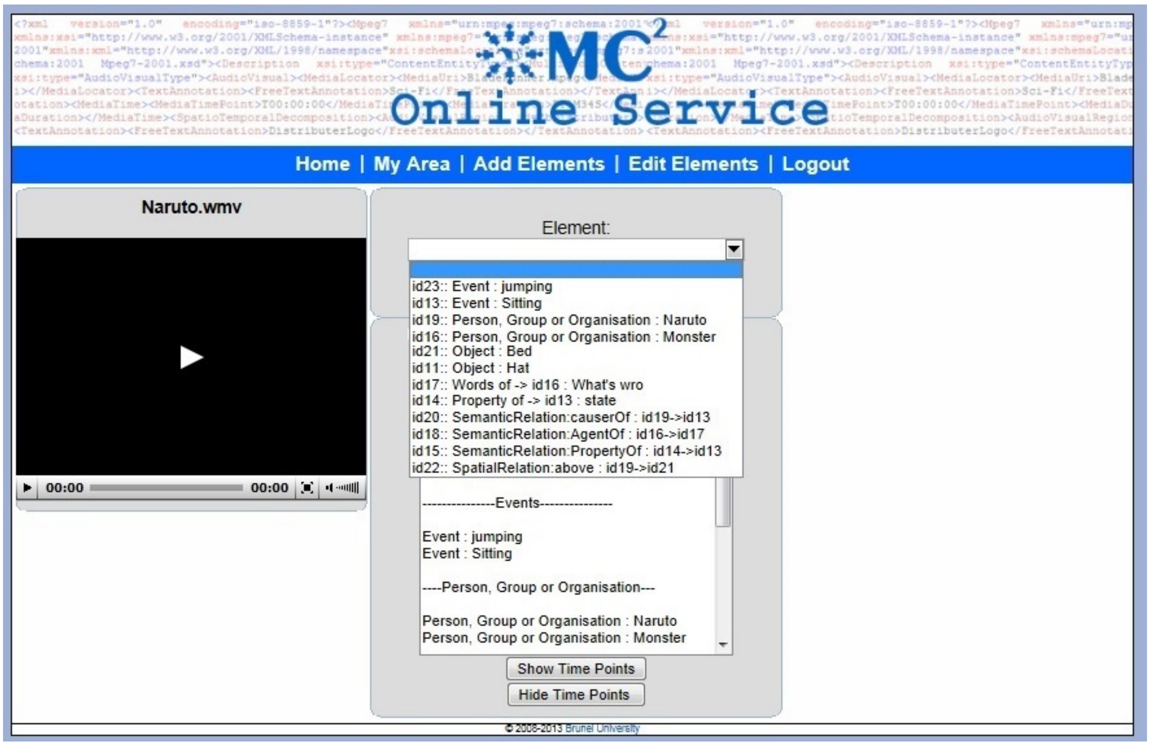


Figure 4.29: Elements listed for Editing

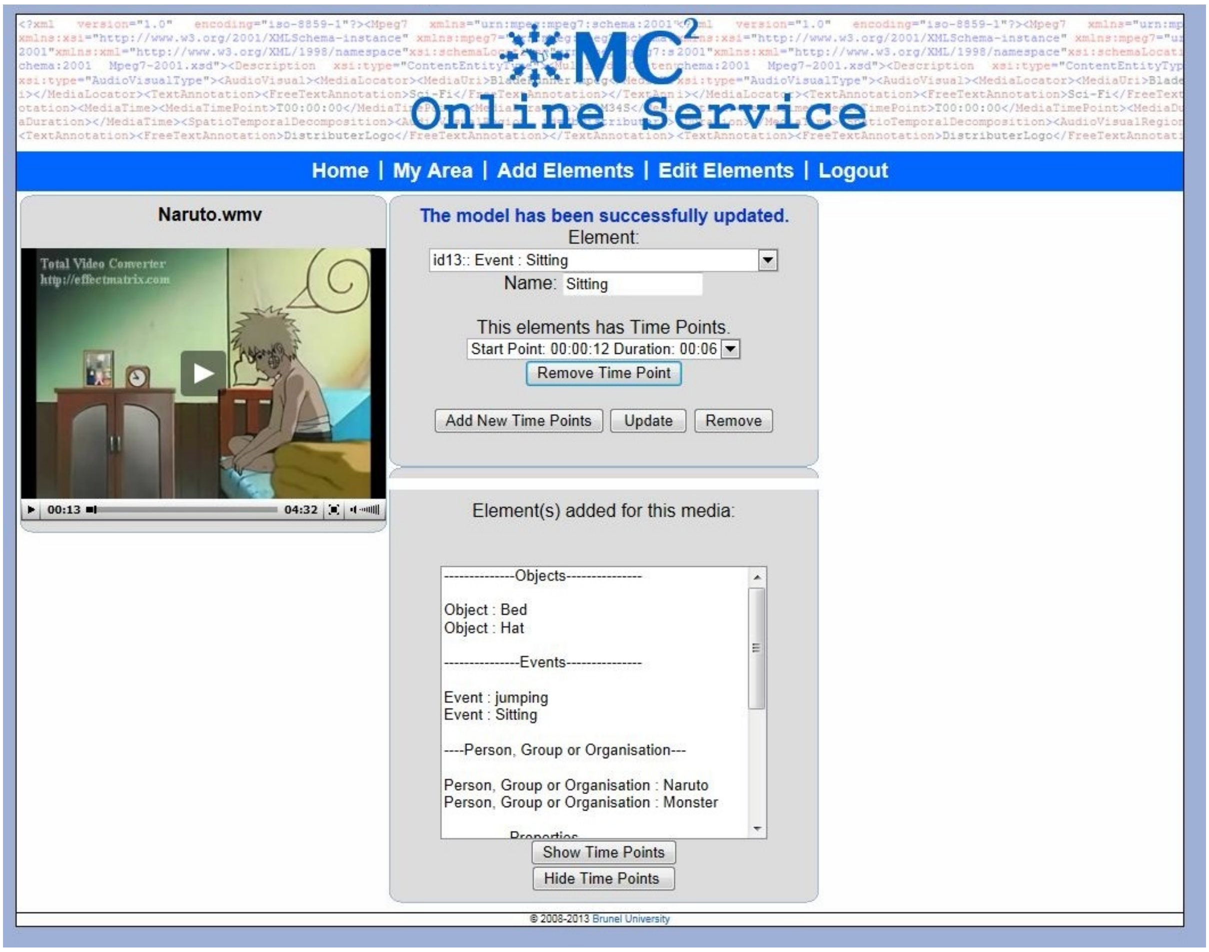


Figure 4.30: johnSmith edits an event added by ikBrunel

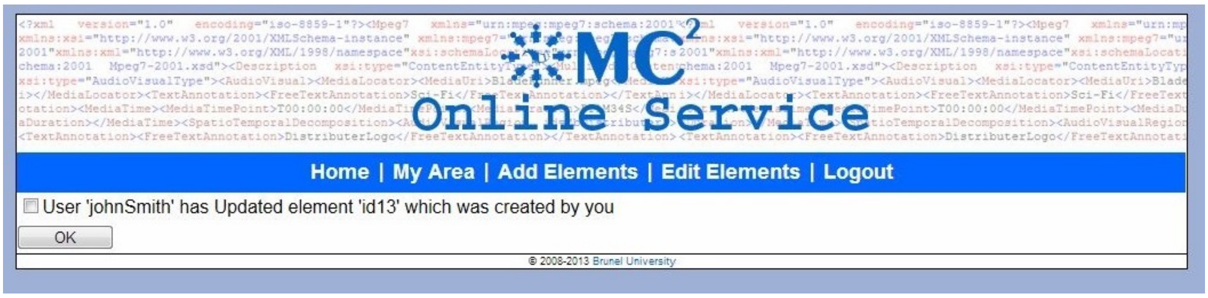


Figure 4.31: ikBrunel receives a notification of update by johnSmith

4.1.15 Collaboration: Commenting, Discussing and Updating

Figure 4.32 illustrates the commenting process. For instance, johnSmith may identify what they perceive to be an inaccuracy in the model and may decide to notify ikBrunel of this. The comments are submitted using the Commenting component which records them as SemanticState DS. Figure 4.33 depicts johnsmith editing the object “hat” as they described in their comments. Based on response by ikbrunel they may decide to continue having a conversation on the set video. In Figure 4.34, they argue that “Sugegasa”, while it might be the correct name for the object, it might also be a word not known to users and, therefore, might hinder users’ chances of locating the media stream.

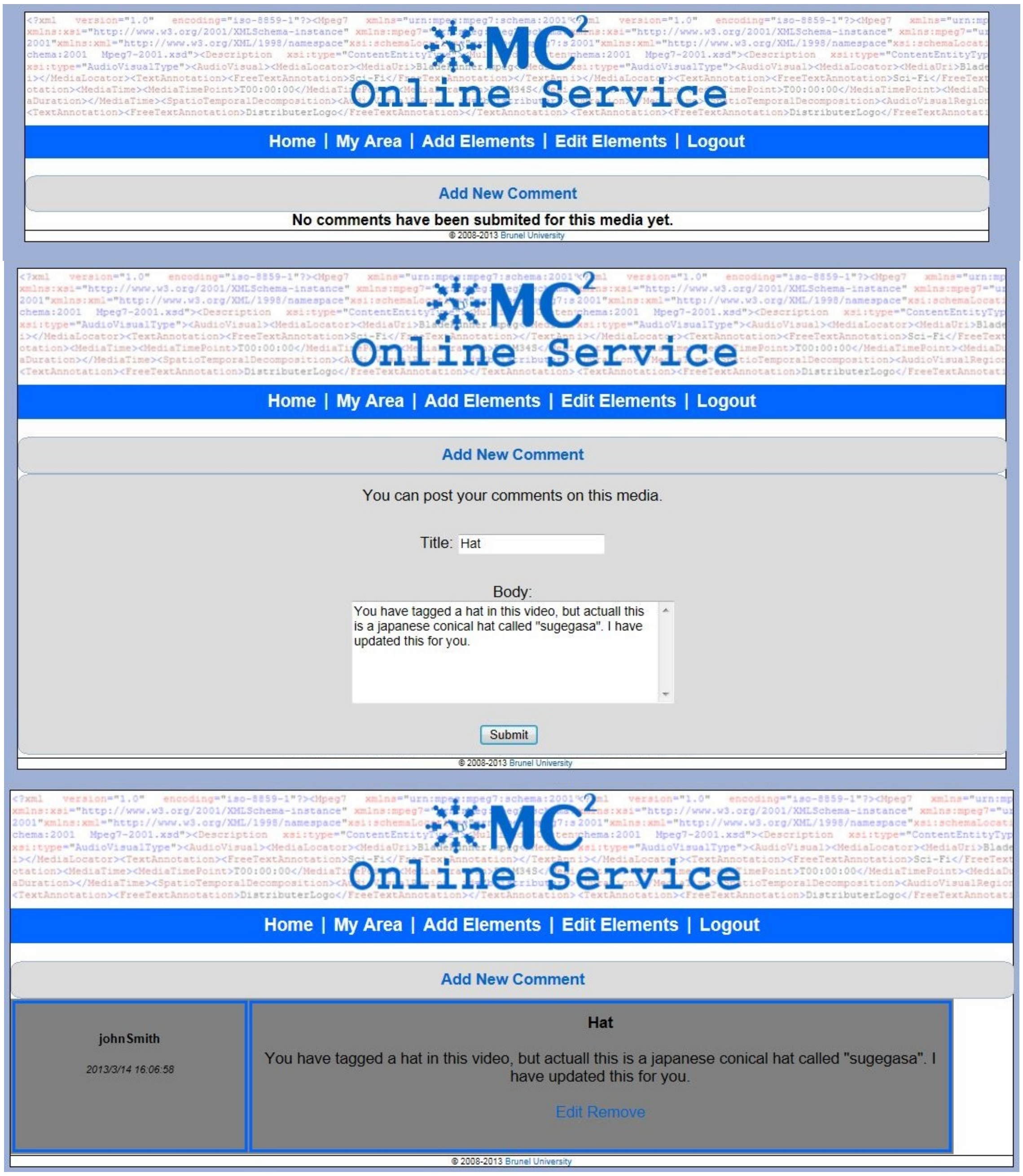


Figure 4.32: Commenting on a video

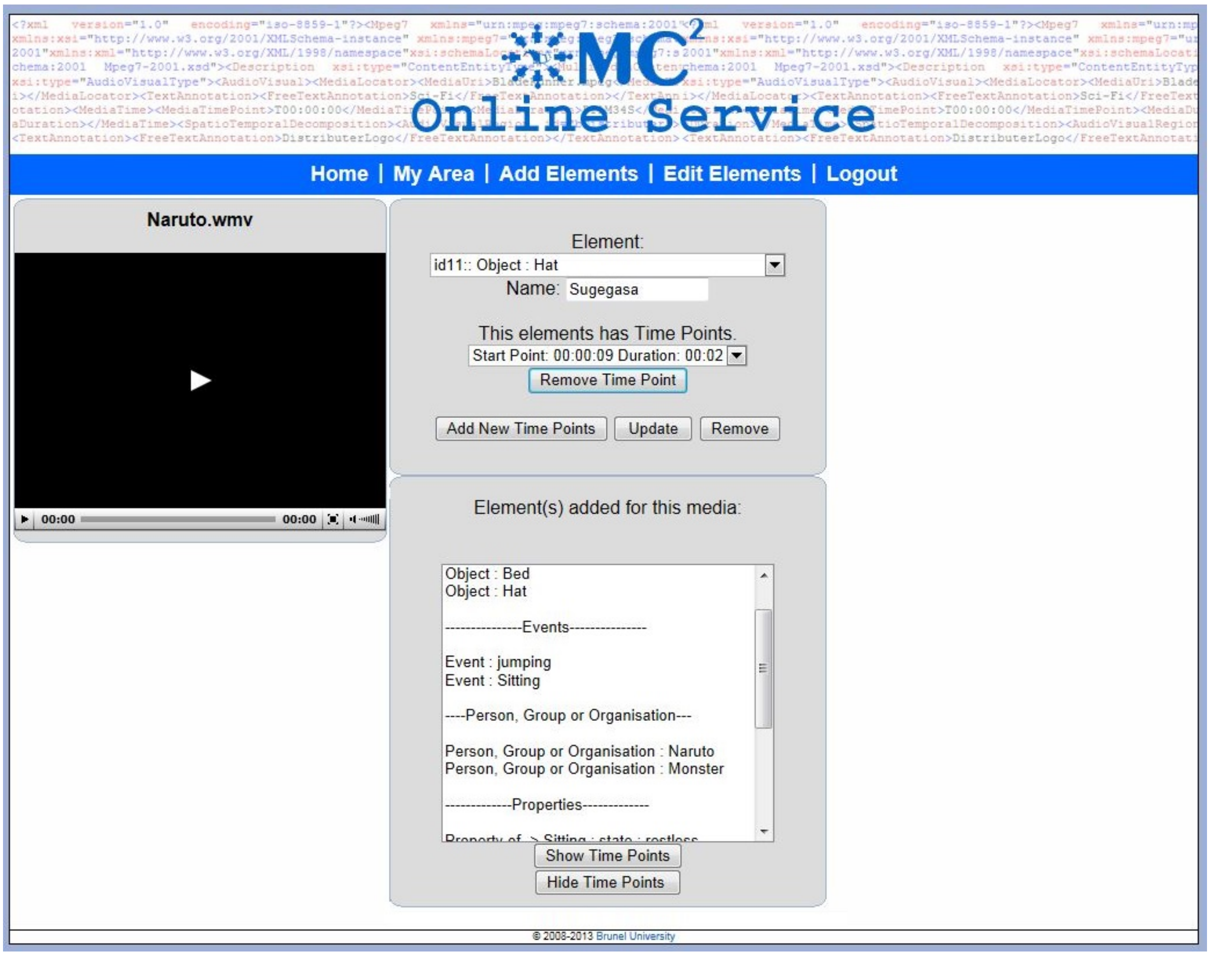


Figure 4.33: Updating the object "hat"

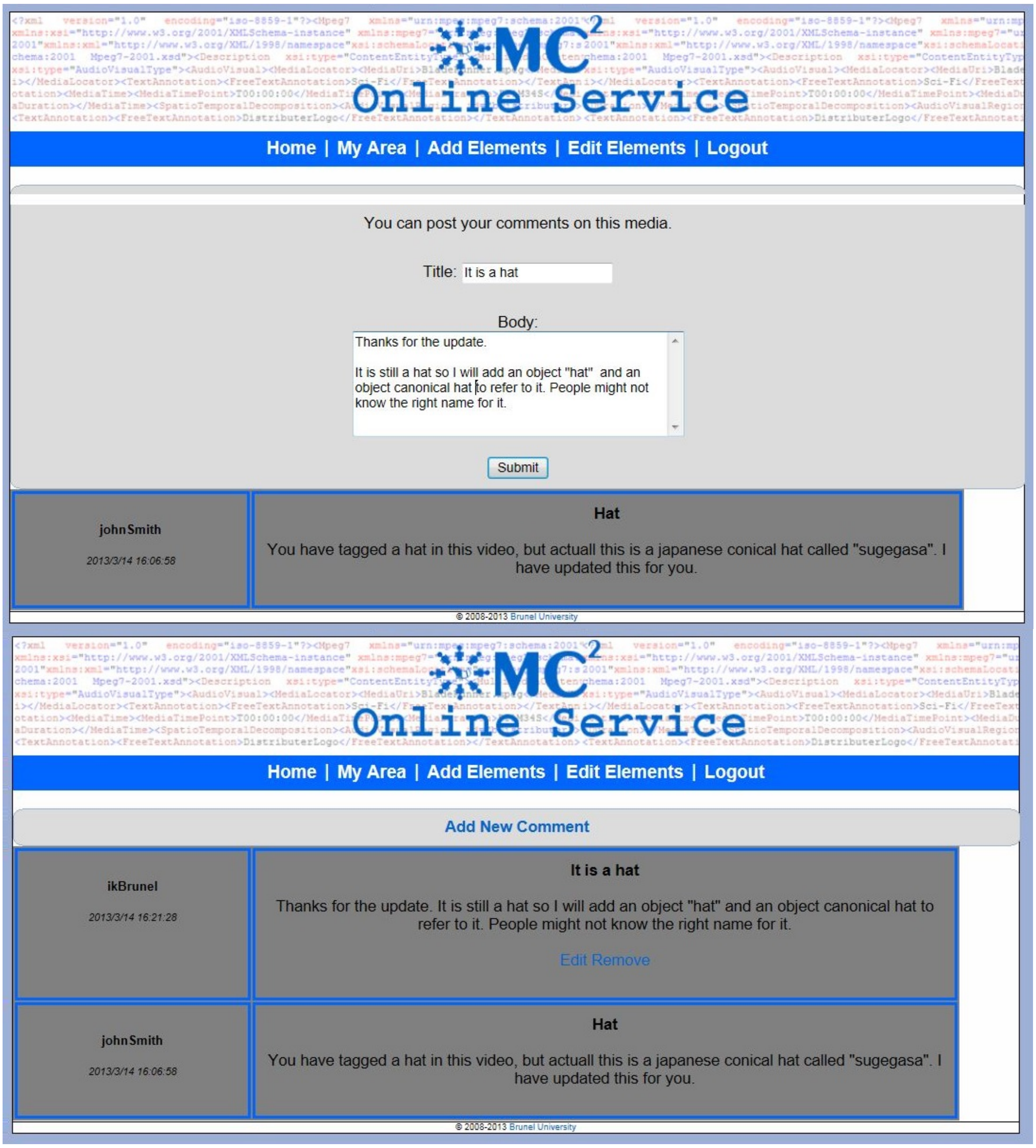


Figure 4.34: ikBrunel replies to johnSmith comment about the object "hat"

Figure 4.35 illustrates ikBrunel's update on the model as per their conversation with johnSmith. In Figure 4.36 johnSmith replies to ikBrunel and suggests that instead of adding a third object called "canonical hat", it is better to add "canonical" as a property of the "hat" object.

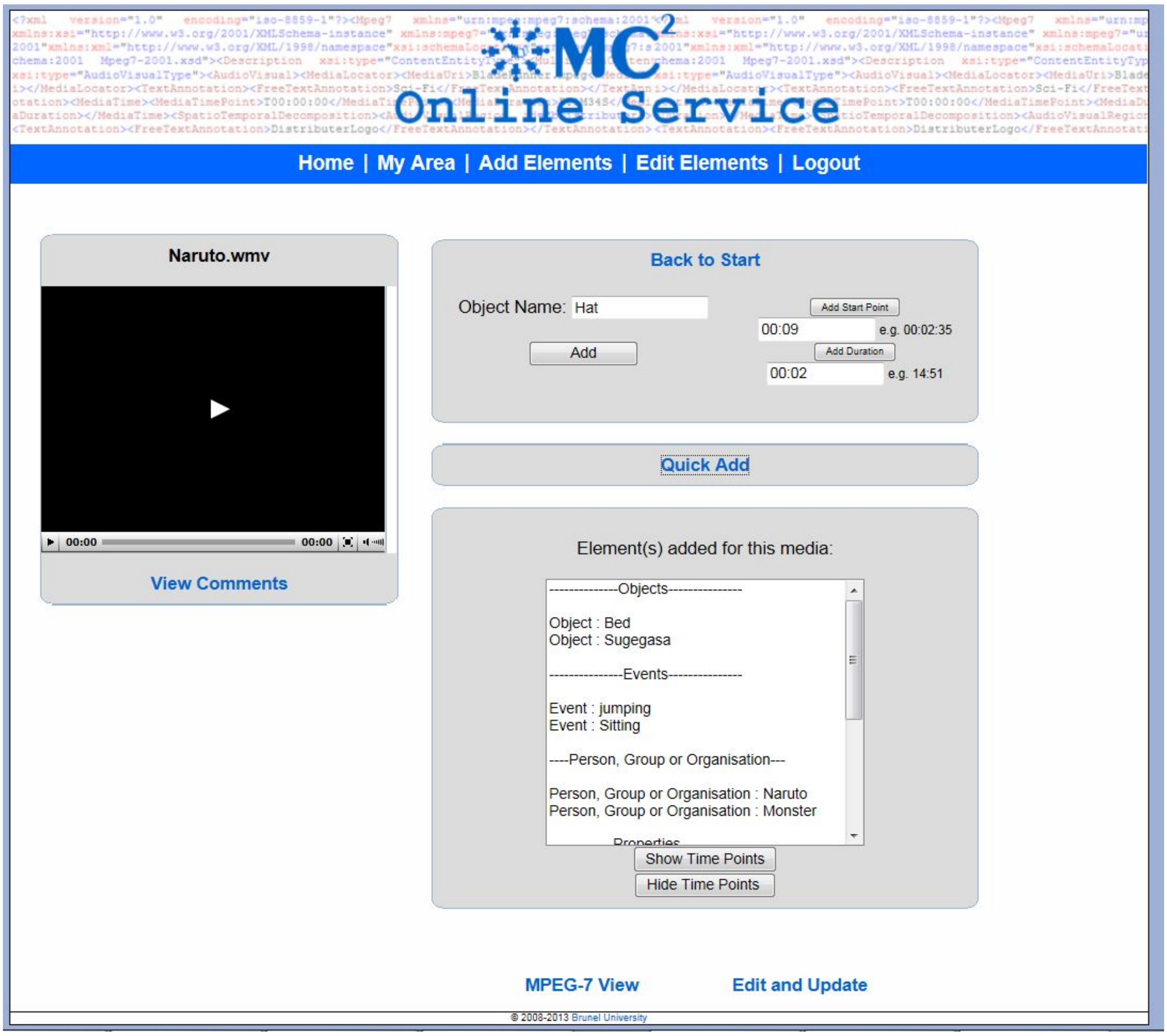


Figure 4.35: Adding object “hat”

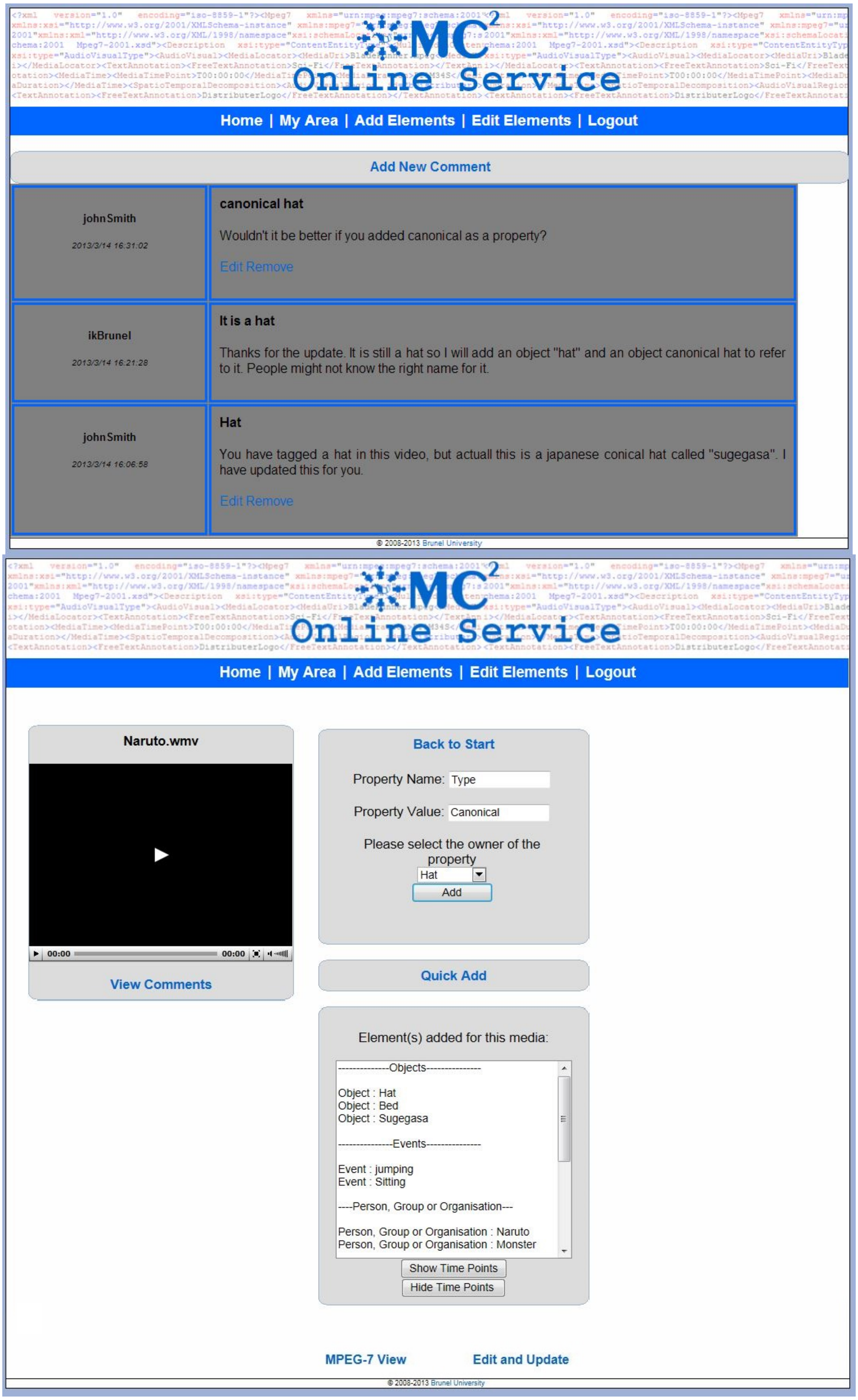


Figure 4.36: Adding property per suggestion

4.1.16 Collaboration: Another User joins in for Tagging

In Figure 4.37 a new user, janeDoe, joins the conversation and suggests that they can define a relation between the objects "hat" and "sugegasa".

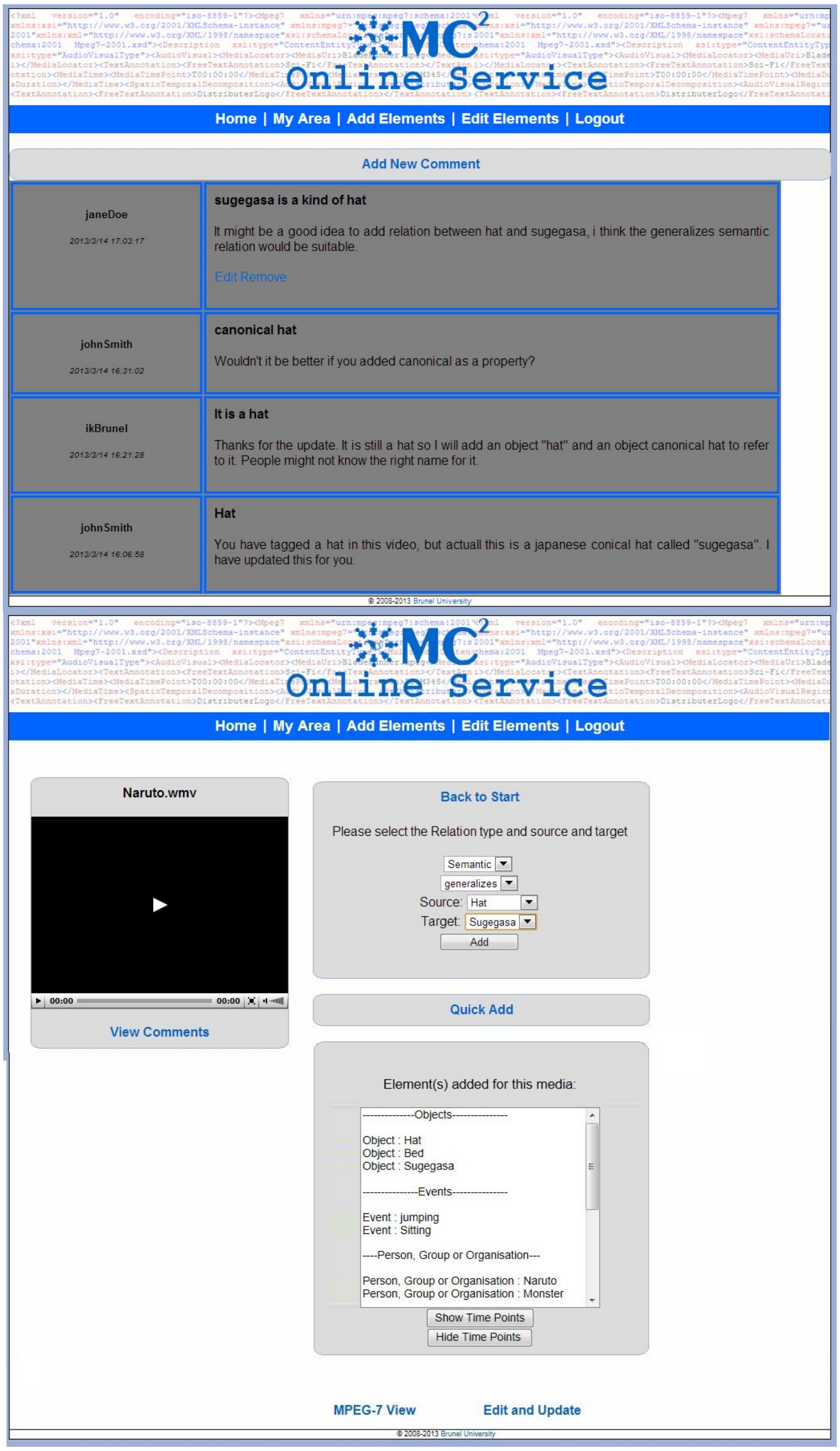


Figure 4.37: janeDoe joins the collaboration and adds a semantic relation

4.1.17 Add Semantic Time and Temporal Relations

In Figure 4.38, janeDoe also inserts a semantic time after she decides that the event "sitting" takes place during the "morning". In Figure 4.39 she connects this semantic time to the "sitting" event as a temporal relation through the Add/Edit Relations component.

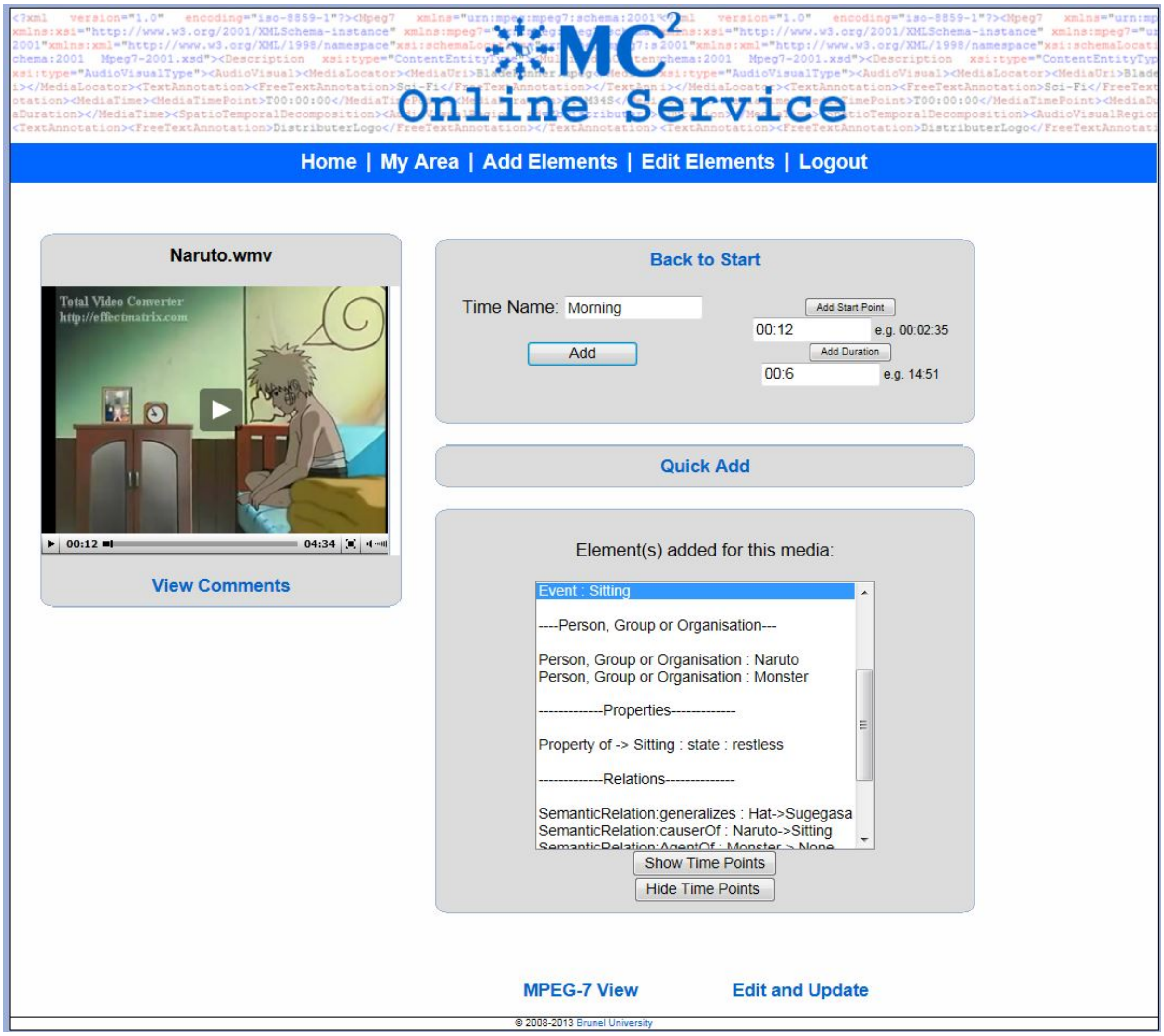


Figure 4.38: Adding "morning" as semantic time

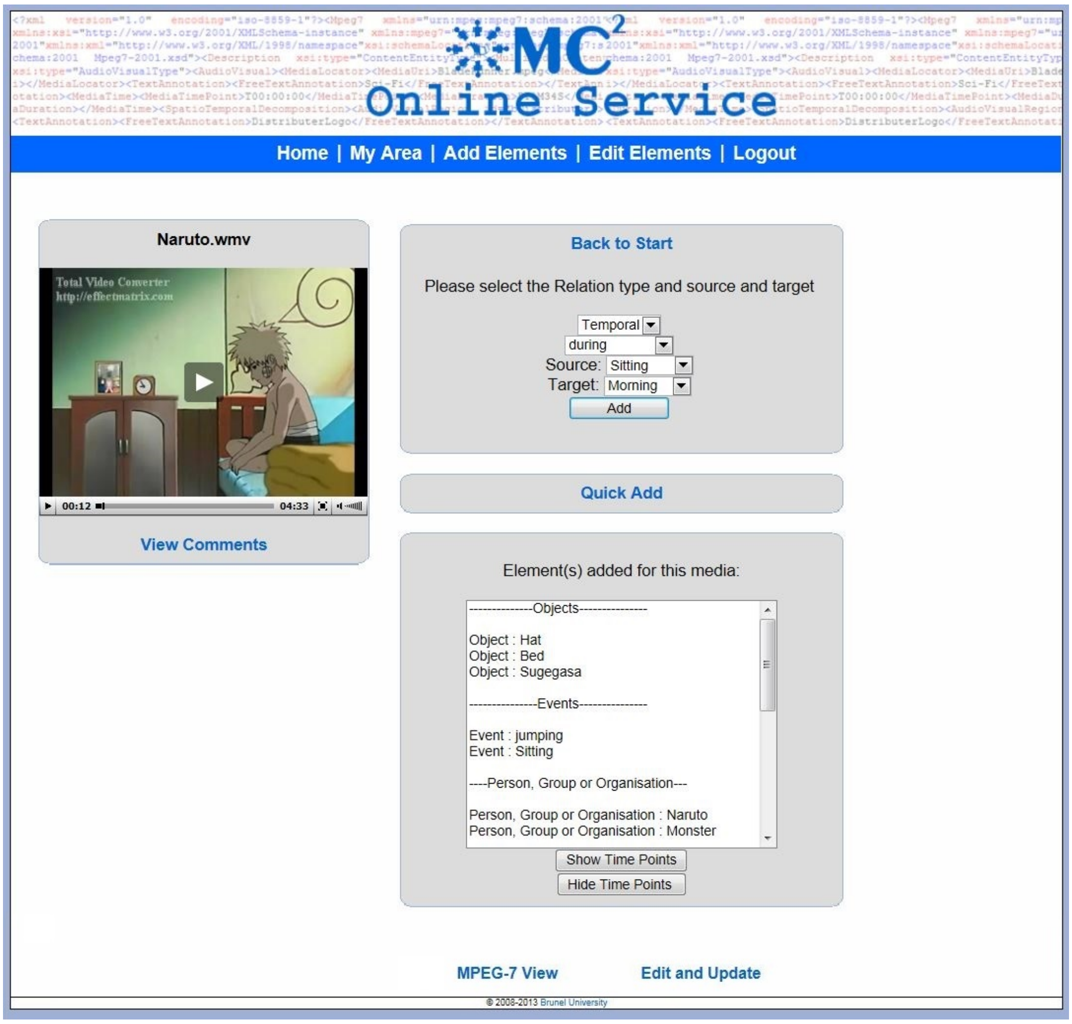


Figure 4.39: Temporal relation added between “sitting” and “morning”

4.2 MC² Validation

MC² validation Is twofold: First, MC² is evaluated as a tool with the same user experiment that led to the conceptual model of user behaviour and with the same groups of users that participated in the initial experiment and then compare the results to those of the initial experiment. Second, the MC² functionality is evaluated by considering its stability, plasticity and accuracy against that of folksonomy and MPEG-7 tools and whether it addresses the implications identified earlier.

4.2.1 Empirical Validation

An experiment was undertaken to evaluate the MC² online service and compare the results with the results of the experiment summarised in chapter 2. The experiment design was kept similar except that only media streams are used, since they incorporate both video and audio and each frame can be considered as a still image. Using the MC^2 online service, the same 51 users from the preliminary experiment were asked to upload and tag a 3-5 minute video (similar to the task they undertook with COSMOSIS, except that they were asked to try and model the frames as still images and the audio of the movie as a music if it was suitable in their opinion). The users were also required to search for other videos on the system and add new tags or update existing tags in them. They were unsupervised, but communicated with other users via the forum and commenting features of the system. The users were afforded six weeks to finish the tasks set for them, which gave them enough time to familiarise themselves with the online service. The users were provided with a service manual (attached as Appendix IV). Users were interviewed after they completed all tasks. Each interview was semi-structured. Grounded theory was then used to analyse the experiment data and the same processes of open coding and memoing were carried out. The full experiment design details, including the detail of the tasks can be seen in Appendix III and interview questions are attached in Appendix V.

MC^2 seeks to balance the positive aspects of both folksonomy and MPEG-7 tools while seeking to overcome their negative aspects. The folksonomy tools provide a collaborative environment with easy to use interfaces that enable users to create metadata free of the limitations imposed by structured metadata. However, this freedom results in confusion, difficulties in searching, semantic ambiguity and identity awareness issues. Conversely, while MPEG-7 tools are very well structured and therefore produce well-formed metadata with precise semantics, which facilitates searching and browsing and raises identity awareness, this structure makes these tools harder to use and creates limitations for the users which often leads to less detailed metadata and complex interface.

4.2.1.1 Tagging

Figure 4.40 compares the tag usage in MC^2, the folksonomy tools that support video (YouTube and del.icio.us) and the MPEG-7 tool (COSMOSIS) as average use of each tag per user in the respective tools (average across all tasks). As the results in the folksonomy tools were very similar, the mean is shown. These results show an overall increase in tag usage in MC^2 while the usage pattern has remained the same. This can be attributed to simplification of the tagging process on the one hand and to the MPEG-7 structure on the other. The figure illustrates that MC^2 has achieved its aim of increasing tag usage through managing the tagging process.

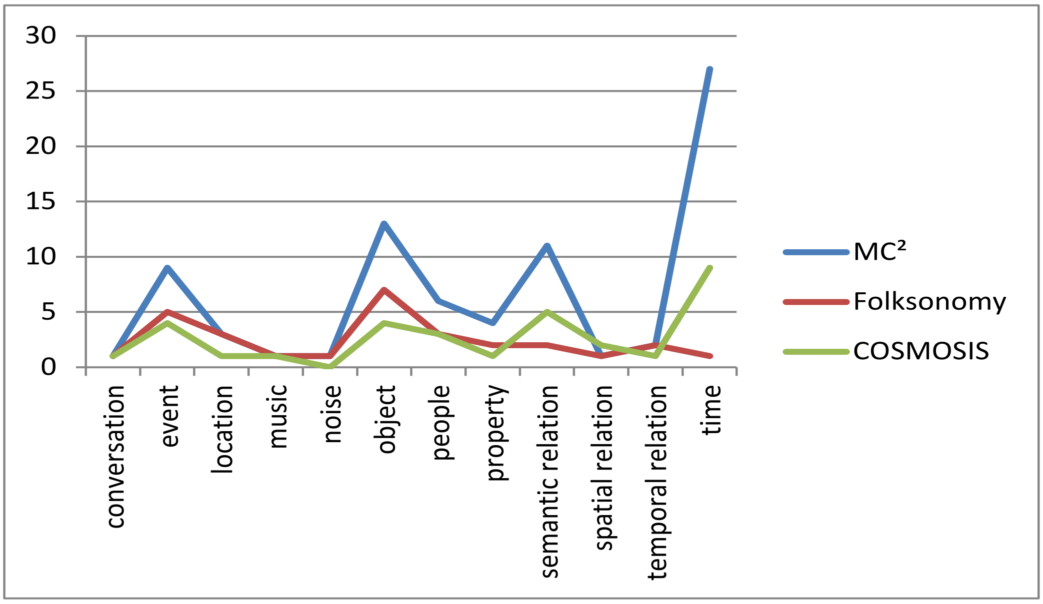


Figure 4.40: Tag usage in MC^2, folksonomy tools and the MPEG-7 tool

Events, objects and people, which were the most used tags in both folksonomy tools and the MPEG-7 tool, are still among the most used tags in MC^2, while time and semantic relations have been used a considerable number of times, which was highlighted by their use in the MPEG-7 tools. The time tag has seen a great increase in use compared to the MPEG-7 tool. This can be explained by the fact that time has been prioritised in MC^2, resulting in better support for this tag in the content model and the interface. This shows that this prioritisation in the framework was met positively by users.

As highlighted in the functional requirements in Section 3.1.2, the framework assigned a higher priority to properties, including higher priority in the content model as well as facilitating the use of it in the interface. This has been positively received by the users, resulting in an increase in the use of property tags.

The results indicate a slight drop in use of spatial relations in comparison to the MPEG-7 tool but a slight increase over the folksonomy tools. As the use of other tags has been simplified and spatial relations were assigned a low priority, this may be explained by the fact that users have tended to use other tags more often and therefore the use of spatial relations has decreased.

4.2.1.2 Context

Figure 4.41 shows the users' views and behaviour towards the tools used in these experiments. The data for this figure was derived from the interviews, communication transcripts in the system and patterns of tag usage.

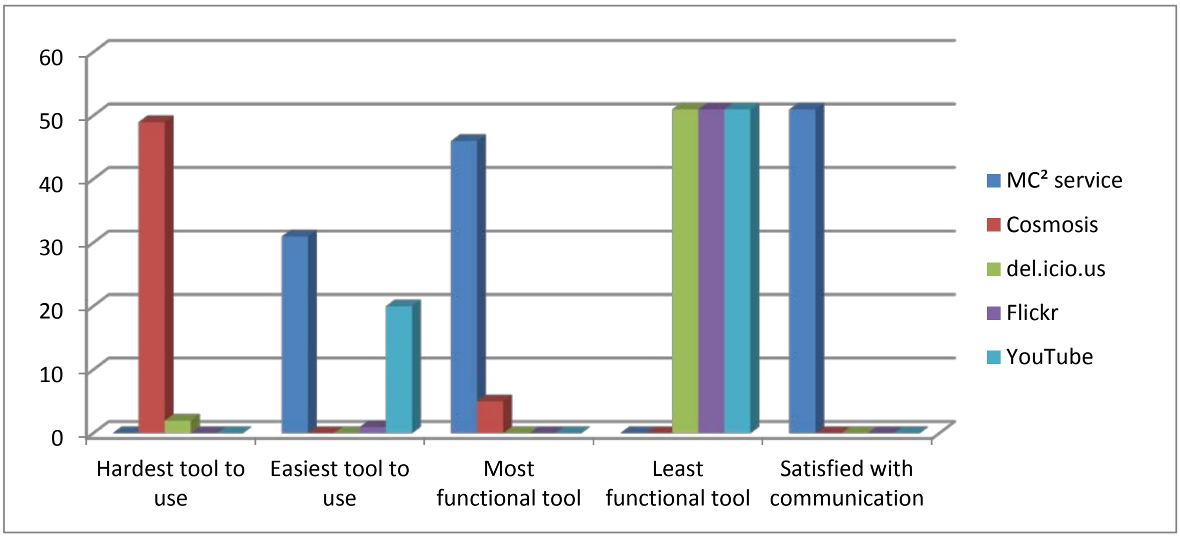


Figure 4.41: Context of tagging systems

COSMOSIS was voted by 96% of the users (49) as the hardest tool to use and none considered MC^2 the hardest tool to use. This suggests that even though both these systems are based on MPEG-7, MC^2 is significantly easier to use. At the same time, 90% of the users (46) found MC^2 the most functional tool, while 5 users voted for COSMOSIS, and none voted for any of the folksonomy tools. This suggests that MC^2 has successfully harnessed the power of MPEG-7 while minimising its complexity such that users not only do not find it hard to use but consider it more functional than COSMOSIS.

The folksonomy tools were together considered the least functional tools by all users. All 51 users claimed that all folksonomy tools used in the experiments were equally low functioning. 31 users (60%) found MC^2 to be the easiest tool to use while the rest of votes went to YouTube (20) and Flickr (1). This majority support for MC^2 suggests that it has managed to provide a comfortable tagging environment for users without sacrificing functionality.

The users in the previous experiment were asked to use live communication through instant messengers while tagging. All the users (51/51) found this distracting and unnecessary. In MC^2, users were asked to use the messaging and commenting capabilities of the service, and all users were unanimously satisfied with this means of communication.

4.2.2 Theoretical Validation

The framework and online service proposed in this thesis aims to meet the implications identified in Chapter 3. Therefore, these implications were considered as criteria by which to evaluate MC^2 and compare these results with those of the folksonomy and MPEG-7 tools used in the preliminary experiment. The implications are categorised under the same criteria as described in section 1.3. Table 4.1 summarises the results of this comparison.

4.2.2.1 Model stability

The implications in this area concern the stability of the model against change. The MPEG-7 profile used in this framework is a structured model with an extensible schema ensuring that even in light of new tags the MC^2 model stays stable since the addition of new tags can be incorporated without modification to the schema of the existing tags. Consequently, despite its comprehensive nature, the proposed framework does not create unnecessary complexity because of its *inclusive and concise metadata structure*. Based on this implication no tag is left out of the metadata while the structural integrity is not sacrificed to achieve this. Since the model is inclusive and concise and the most important tags discovered previously in the experiment are *prioritised into metadata core*, the metadata model is stable against adding new tags. Folksonomy tools, such as those used in the experiments, are inherently unstructured. While this enables inclusivity, it sacrifices conciseness and structure. This lack of structure and emphasis on complete freedom in tagging is at loggerheads with the establishment of a metadata core and tag prioritisation.

The ability to model and search on semantic concepts of time such as “evening” or “Middle Ages” which are also comprehensively connected to related objects, events, properties and relationships, *facilitates semantic time for all features*. For example connecting properties for the previous examples could be “warm evening” or “Middle Ages: 400 AD-1400 AD”. Similarly, connecting these semantic time tags with events such as “riding a bike in a warm evening” allows the users to create more complete, interconnected models. While MC^2 fully addresses this, COSMOSIS only partially does because, in spite of allowing for semantic time to be added, it does not do so for all features. This results in the metadata lacking the inclusiveness and conciseness these implications require.

Another pair of implications that have been accommodated in MC^2 and that contribute to model stability are those relating to the collaborative nature of this model. MC^2 proposes an extensive profiling system that logs every action taken by the user. By *facilitating showing all tags and authors per content* allows for management which counters rudimentary tag-based access and anonymous tag authorship. Consequently, it is very clear which tags have been created and edited by which users. This expansively increases stability of the model for users. In the folksonomy tools, such as YouTube and Flickr, it is not possible for other users to contribute tags and the user is never made aware of any modifications to the tags he/she himself/herself has made. Another problem with these systems is that the tags are not defined separately and consequently the process of adding and editing tags is completely hidden to the user. COSMOSIS in itself is not a collaborative system and therefore there is only one author for each model. The MC^2 framework also provides communication between users through commenting and messaging due to *emphasis on asynchronous communication*. This enables users to communicate with each other and establishes the identity of users within the system.

4.2.2.2 Model plasticity

Model plasticity concerns the degree by which the model stabilises after a finite number of metadata additions so that no further model elements are required in order to achieve full closure of the semantic gap by the MPEG-7-based content model. The framework proposed here ensures that once all the DSs in the scheme are used in the model, the user will be able to tag anything that is in the media file without having to use a new DS. Using all tags available in the model along with extensive profiling capabilities through *facilitating showing all tags and authors per content* allows for a placid model as the user can see which tags are missing from the model by reviewing the existing stage of the model. It also allows the users to propagate tags through allowing tagging of sets of streams. COSMOSIS partially addresses this since a tag can be propagated to all frames in video, used tags may be assigned to new videos, and it shows users the set of tags added for a video.

As opposed to folksonomic tools, MC^2 allows the user to define their own relationships between the tags they have defined, *facilitating connection between co-related tags*. For example, the framework enables the user to add the object “bicycle” and event “riding” and then connect them together to create “riding a bicycle”, while this event can also be connected to “horse” creating the semantic concept of “riding a horse”. As users are encouraged to actively associate properties with other content features, the MC^2 framework provides capabilities to *facilitate property description*. Therefore, the users are able to model more complex concepts such as “riding a red bicycle” where “red” is a property of “bicycle” hence closing the gap even further. COSMOSIS partially facilitates connection between co-related tags as it allows the definition of relations between certain tags. MC^2 also provides the means to tag sets of media streams as opposed to carrying this out sequentially, a facility not provided in existing systems. Therefore, the user can propagate metadata through to a group of media streams quite readily, as opposed to just adding them to each one separately.

MC^2 uses WordNet to control the tags that are entering the system and suggest the synonyms to the user so that they can be added as well. In the same way, WordNet also suggests connected terms that relate to the tag added by the user in order to *enforce ease of use,* therefore mistakes and misspellings are avoided. The existing folksonomy systems and COSMOSIS do not provide any means for synonym control; users are free to tag as they like.

The *quick add facility* enables users to quickly add more tags trying to complete the semantic gap. Finally, *free-text tagging within the interface* allows the users to freely describe the semantic features of the content that the user believes are not encompassed by any existing tags, and subsequently connecting them to existing tags.

4.2.2.3 Model accuracy

Implications here focus on how accurately the model relates to real world semantics. A comprehensive and concise model should include all possible semantic features, because such a model will be the most accurate as it has an *inclusive and concise metadata structure*. Unlike the unstructured text of many existing tools, the structured tagging of MC^2 aims to clarify the tags through the MPEG-7 structure, which illuminates what type of tag each keyword stands for (e.g. object, event, time). It will also increase the expressiveness of the model by defining the relationships between the tags. For example users will not have to tag two events of “riding a horse” and “riding a bicycle” as they can create the event “riding” and connect it to objects “horse” and “bicycle” which is much more accurate. This structured tagging also serves to increase the accuracy of the search results as the results may be ranked according to the usage within the content models. Some folksonomy tools (del.icio.us) have clouds that represent the demographics of the tags in media but this information is not used in ranking the media and searches, therefore, these clouds cannot be considered to meet this challenge.

In conjunction with a concise and inclusive metadata, MC^2 allows users to tag semantic time features. By *facilitating semantic time for all features*, MC^2 ensures for model accuracy through allowing users to create semantic time features such as “middle ages” while defining it as “400 AD to

1400 AD". This adds to the accuracy of the model because users can find middle ages even if they don't know the date range, or they can find the name through the date range.

COSMOSIS and MC^2 both allow the user to add time points to the tags in the form of starting point and duration which helps to focus the attention to a specific time during a media stream by *facilitating addition of time points*. Yet folksonomic tools do not provide such facility and, therefore, it is complicated and difficult to focus the attention of all users to the certain tags at the required time during the video. For example, in the MC^2 model an event "riding" could be defined to have happened exactly at 2 minutes and 22 seconds through the video, lasting for 15 seconds. YouTube allows visual commenting on the videos with embedded time points but this is not the same as adding time points to tags as they are not related to any tags.

MC^2 ensures a high level of accuracy in the model by *enforcing ease of use* using WordNet and tag suggestions. The WordNet dictionary is used to identify and notify the user of misspellings. MC^2 also provides a strong model accuracy through its versioning component which not only keeps track of all the actions users perform, but also notifies the original creators of videos and tags of new tags being added or tags which they had created which have been amended or removed. *Facilitating showing all tags and authors per content* will clarify when and by whom changes are made and it will always notify the owner of content of any changes to the model. This means that changes to the model can be controlled, monitored and if needs be corrected.

Table 4.1: Comparing MC^2 with existing tools

Implication	**Criteria**			**MC^2**	**Folksonomy tools**	**COSMOSIS**
	Stability	***Plasticity***	***Accuracy***			
Facilitate semantic time for all features	✓		✓	F	-	P
Facilitate connection between co-related tags		✓		F	-	P
Prioritise into metadata core	✓			F	-	P
Inclusive and concise metadata structure	✓		✓	F	-	P
Facilitate property description		✓		F	-	P
Facilitate addition of time points			✓	F	-	F
Enforce ease of use		✓	✓	F	P	-
Facilitate showing all tags and authors per content	✓	✓	✓	F	-	-
Emphasis on asynchronous communication	✓			F	F	-
Quick add facility		✓		F	-	-
Free-text tagging within the interface		✓		F	F	-

F = full support   P = partial support   - = no support

4.3 Summary

The chapter maps the road to the validated MC^2 framework. First, a walkthrough of the MC^2 online service demonstrates a typical "day in the life of MC^2. Secondly, a two-part validation evaluates the service both as a user tool and its functionality against folksonomy and MPEG-7 tools. The two-part validation produces the expected results with respect to the implications that arise from the shortcomings of folksonomy and MPEG-7 tools. The following chapter concludes the thesis with a summary, research contributions and future research and development.

Chapter 5: Concluding Discussion

This chapter gives a summary of the thesis, the research contributions made and makes suggestions for future research and development. It revisits the motivation behind the research undertaken, and reemphasises the novelties underlining the three contribution; MC^2 framework, The conceptual model of user behaviour and the MC^2 online service.

5.1 Thesis Summary

Chapter 1 examines the effectiveness of existing approaches for multimedia tagging tools and categorises them based on tagging (structured or unstructured) and context (individual or collaborative). The chapter also unravels future challenges with collaborative content modelling in three categories, namely, model stability, plasticity and accuracy. The chapter then draws its research aim and objectives and the research methods used in pursuing each objective.

Chapter 2 derives a conceptual model of user behaviour as a fuzzy cognitive map. This fuzzy cognitive model describes the relationship between users' tagging behaviour and context. Data collection for model building is made through a series of experiments with existing multimedia tagging tools and a group of users. Data analysis both of the experiment results and user interview transcripts is made using grounded theory.

Chapter 3 designs and develops MC^2, an MPEG-7 based framework for collaborative content modelling based on the fuzzy cognitive model derived in chapter 2. The set of implications associated with the fuzzy cognitive model served as the basis for the development of the framework. The chapter implements an online service based on MC^2 using rapid application prototyping to serve both as a proof of concept and as tool for the participants and any interested parties.

Chapter 4 validates the MC^2 framework through its proof of concept. The chapter first offers a walkthrough the MC^2 online service demonstrating its capabilities and collaborative nature. The chapter undertakes a two-part validation process of the service. During the first part, the online service is put through the same experiment as that described in chapter 2 using the same group of users and the results are compared with those of the initial experiment. During the second part, the online service functionality is evaluated against that of folksonomy and MPEG-7 tools in terms of stability, plasticity and accuracy and the set of implications discussed in chapter 2.

5.2 Research Contributions

The thesis makes three research contributions which are listed in order of priority in Table 5.1 against the relevant research objective.

Table 5.1: Research contributions and objectives

Contribution	Objective
MC^2 framework	O2: Design and develop a framework for MPEG-7 content modelling communities based on the conceptual model of user behaviour
Conceptual model of user behaviour	O1: Design a user experiment in order to study users' tagging behaviour with existing multimedia tagging tools and identify any relationships between such user behaviour
MC^2 online service	O3: Implement the online service based on the MC^2 framework developed as a proof of concept

5.2.1 The MC^2 Framework

The MC^2 framework is the core contribution of this research and has been designed and developed in pursuit of objective 2 to cater for collaborative multimedia content modelling using the MPEG-7 standard. The framework fully addresses the set of implications for multimedia content modelling communities unravelled in Chapter 2. Hence, it fills the gap for a comprehensive framework that is based on the needs and behaviour of the users and produces comprehensive content models. MC^2 consists of independent components allowing it to be customisable.

At the heart of the framework lies an MPEG-7 profile designed to support and facilitate collaborative tagging while it also caters for the needs of a collaborative community. As well as a comprehensive content model, the MPEG-7 profile also includes user profiles, user action history and archiving/versioning models. To allow for a comprehensive MPEG-7 content model that is customised to match user behaviour, the framework prioritises the fundamental tag elements that were discovered in the empirical research, reported in Chapter 2, into the metadata core. While the framework prioritises some tags (i.e. objects, events, people) it incorporates all possible tags as well as creating the facilities to add tags that are not specifically supported by MPEG-7. The framework also enables modelling of conversations which is not specifically implemented in MPEG-7. The framework enables users to tag conversations through using SemanticState DS and agentOf semantic relation for describing a conversation and connecting it to the speaker respectively. Considering that the framework allows adding time points to all tags including conversation, this facility in the framework provides the potential for new research into subtitling videos. The incorporation of user profile and user action history in the framework provides the grounds for increasing the identity awareness amongst the users of the community which not only caters for collaborative contribution but also encourages it. The framework stores every modelling action of the user into the user model, which enables users to track their interactions as well as versioning.

The archives model, which is also part of the MPEG-7 profile of the framework, stores all actions on content models, ensuring that the framework can demonstrate all the tags and author for each element of content, while being able to keep track of all changes that are updated to the content model. The asynchronous communication through notifications and commenting, incorporated within the framework, ensures that the collaborative nature of the community and the need for communication amongst contributing users are preserved without creating distraction to the tagging activities. The framework also assures for accuracy and comprehensiveness of the content model by incorporating recommendations after tagging, providing preventive measures against mistakes and misspelling, and measures for connecting synonym tags.

5.2.2 The conceptual model of user behaviour

The fuzzy conceptual model of user behaviour is the first contribution of this research and has been derived in pursuit of objective 1 through a user behaviour experiment with existing multimedia tagging tools, namely YouTube, Flickr, del.icio.us and COSMOSIS, and grounded theory analysis. Since there is hardly any empirical data on how users tag multimedia content, the conceptual model fills a gap for mapping user tagging behaviour. The conceptual model should provide new insight for researchers into understanding how collaborative user behaviour unfolds.

The conceptual model shows that while users' tagging behaviour is dependent on the functional availability and feature frequency in the content, the user feature preference still plays the most significant role in tagging. This was backed up by the comments in the interview were participants commented that if certain features were available, they would have modelled more tags. The user behaviour shows that users have functional preferences that are not fully included in existing tools. The users would like to be able view all the tags added to a media stream and also prefer to be able to see who has added these, hence, the need to facilitate showing all tags and authors per content. The experiment clearly indicates the need to reach a balance between the metadata structure and its complexity which is reflected in MC^2.

Surprisingly, almost all users argue that while some sort of communication is essential in collaborative tagging, they find live communication extremely distracting and go on to emphasise the need for asynchronous communication whilst engaged in collaborative tagging.

5.2.3 The MC^2 online service

The MC^2 online service is the third contribution of this research and has been implemented both as a proof of concept of the MC^2 framework and as a community tool in pursuit of objective 3 using rapid application prototyping. The service is validated in Chapter 4 using the same experiment and group of users as in Chapter 2 and the results demonstrate higher tag usage than the folksonomy and MPEG-7 tools deployed in the experiment. The validation also shows that the service functionality enabled ease of use in comparison to the other tools. Through the MC^2 online service, the thesis demonstrates that the MC^2 framework has managed to strike a balance between content tagging comprehensiveness and user tagging behaviour. The online service is now available for public use by any interested party.

5.3 Research Limitations

The first set of limitations focuses on the users and their involvement in the experiment. For such a large-scale generalisation of this research‘s results, the group of participants, 51, taking part in the experiments is relatively small. There are no guidelines on how many participants are sufficient to generalise the results of such an experiment. For every additional participant, the additional time management, data collection and analysis grew exponentially.

Using even less participants but monitoring them for a period of time whilst engaging in tagging activities in public, may provide an in depth insight into user behaviour and preferences and may also reveal trends which has not been uncovered during the experiments. Monitoring users allows the researcher to identify states of confusion, excitement and frustration and when later questioned on those states to complement the insight drawn into their behaviour and preferences.

The users may have been able to provide alternative insights to the design of the framework and the system if they were given the opportunity to pay input in the analysis and design processes. Their diverse backgrounds, level of knowledge and competence with the research pursued and the time factor contributed heavily against such an involvement.

Another limitation lies with the number of folksonomy tools used in the experiment. There is countless number of such tools, many of which have been cited in this thesis. However, not all of these tools support tagging multimedia content and likewise considering all the existing multimedia folksonomy tools during the experiment would be unmanageable let alone collate usage data from each one. Many of these tools, whilst in the public domain, are under-utilised, hence, their statistical significance would be unfairly elevated as a result of inclusion within the experiment. The tools chosen for the experiment were among the most used for multimedia tagging.

Grounded theory, whilst an impartial and unbiased analysis method that allows qualitative analysis of quantitative data, also proved to be a limitation. It is complex and time consuming because of its coding and memoing processes which usually affects the test size of the population. Furthermore, its' increased flexibility brings methodological limitations and, hence, a heavy reliance on the researcher's knowledge and skills.

Finally, the research neither does evaluate the precision of the tags created in the two experiments nor does it perform any recall evaluation on the quality of the tags added.

5.4 Future Research and Development

Research and development that is either underway or planned since completion of MC^2 is illustrated in Figure 5.1.

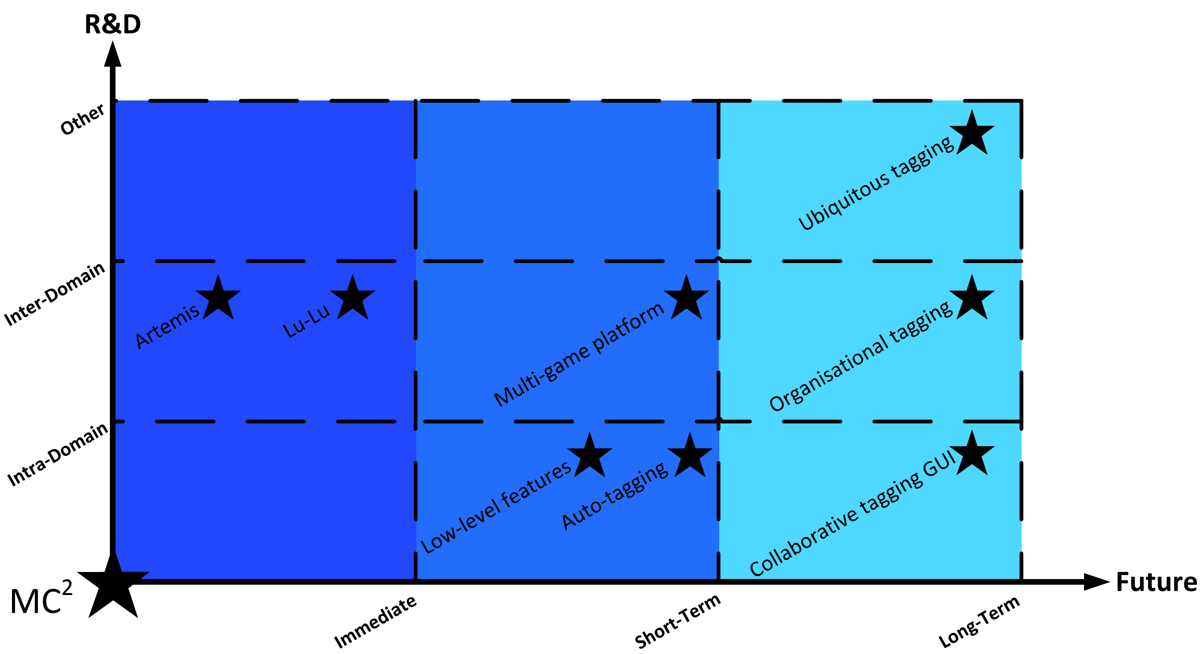


Figure 5.1: Future research and development grid since MC^2

5.4.1 Immediate Future

Over the last decade, the rising popularity of Massively Multiplayer Online Games (MMOGs) such as World of Warcraft (Shirmohammadi and Claypool, 2009) has resulted in a shift towards highly interactive collaborative gaming environments in which the personal needs and preferences of users are preserved. Most of the MC^2 framework has been deployed in game personalisation of MMOGs (Daylamani Zad et al., 2012). The development of Artemis (Daylamani Zad et al., 2012) demonstrates the application of MC^2 in modelling the game content and the player preferences, behaviour and devices in order to support player personalisation and collaborative action.

Collaborative decision making (Williams et al., 2010; Green et al., 2010; Singh et al., 2012) in MMOGs has transformed their scope from merely Ludic, to hoist entertainment, to Lusory, to serve as a means to an end. The deployment of the MC^2 framework in Lu-Lu, an MMOG whose scope is collaborative decision making (Daylamani Zad et al., (Forthcoming)) demonstrates that modelling game content, player profiles, actions and decisions during play is entertaining and supports and encourages players to engage in collaborative decision making.

5.4.2 Short-Term Future

The framework focus is on content semantics, yet expert users might want to tag low-level metadata such as object coordinates in a frame (Tsingalis et al., 2012; Bailer et al., 2012; Ayad et al., 2012). MPEG-7 provides description tools for low-level features which can be included in the content model. These low-level features will add to the accuracy and inclusiveness of the metadata but they would make the tagging process very complex. In order to preserve the ease of use and provide such facilities, the interface would need to be adapted so that users from various levels of competency can access the facilities they need without complicating the interface. Expansion of the framework in order to accommodate low-level content metadata for expert users is under consideration.

The inclusion of object detection functionality (Tan and Lukose, 2012; Myung Jin et al., 2012; Shotton et al., 2013) can ease the tagging process for the users so that once a user has tagged an object, this can trace the object throughout a video and automatically tag it. The framework can be easily endowed with such an automated tagging functionality in order to ease the tediousness of the tagging process. Alongside that, the recommender functionality can also be extended to include automatic searching and addition of tag synonyms to the content model.

The development of Artemis (Daylamani Zad et al., 2012) demonstrates that personalisation within a single game increases game enjoyment and in turn player loyalty. However, the level of recommendations may be inappropriate for a seasoned player joining the game from another game of the same genre. For example, a seasoned player of “First Person Shooter” games would command recommendations at their competence level when switching to a new shooter game rather than at the novice level. Studies on multi-game user behaviour (Grimm and Mengel, 2012; Huck et al., 2011) suggest that developing multi-game platforms where the same degree of personalisation currently offered within a single game is offered across a multitude of games within a genre might have a positive impact on player satisfaction if their skill levels are recognised when switching from one game to another of the same genre. Games within the same genre usually share many characteristics such as game semantics and environment, therefore, a player’s goal, skills and behaviour would be very similar across games of the same genre. Hence, a comprehensive player profile initially created for a game within a genre can be deployed across the entire genre, eventually becoming a player’s genre profile. The main challenge would be in identifying those characteristics that define a genre and the relevant player information that needs to be recognised and modelled in such a profile. Artemis was initially developed with this in mind (Daylamani Zad et al., 2012) and whilst personalisation currently spans a single game, it is possible to extend Artemis’ player profiles to include a multitude of games belonging to the same genre.

5.4.3 Long-Term Future

The conceptual model of user behaviour levies much emphasis on collaborative tagging in order to achieve a balance between the metadata comprehensiveness, structure and complexity and the effectiveness of the tagging GUI. Several studies suggest that a collaborative tagging GUI may aid with collaborative tagging, on one hand, and provide better understanding of user behaviour during collaborative tagging, on another (Yuill and Rogers, 2012; Carvalho et al., 2012). Expanding the MC^2 interface to accommodate collaborative tagging and setting up an appropriate user experiment to investigate such collaborative user behaviour promises to shed new light on the conceptual model of user behaviour, which in turn may inform the development of better GUIs for such collaborative content tagging activities.

Collaborative tagging may be used for information retrieval within organisations to overcome information overload issues (Hsieh et al., 2009; Arnold et al., 2012; Jackson and Smith, 2012). Digital information in the format of documents, emails and attachments, charts and digital conversations can be treated as content, stored in a repository, and may, therefore, be tagged in the same way as multimedia content is tagged. However, Jackson and Smith (2012) argue that web-based systems do not allow users to locate local files and as a consequence users would need to upload every single file they would want to retrieve. This requires users to download files which may result in multiple copies across local machines within an organisation. Jackson and Smith (2012) propose a method whereby the repository server can serve as a folder on the local machine and demonstrate that this method not only improves retrieval performance but it also generates greater cost effectiveness. Institutionalising participation in MC^2 through standardisation is currently being considered, but this requires that the MC^2 service can accommodate organisational tagging without the pitfalls identified by Jackson and Smith (2012). The MC^2 service may be used for organisational tagging in order to annotate and connect stored information collaboratively but this would require three changes: first, the user interface to MC^2 needs to accommodate collaborative tagging in the manner discussed at the start of this section, secondly, the viewing subsystem needs to serve as a non-web local interface and, thirdly, the media management subsystem needs to accommodate all digital file types.

For a while, users have been able to browse a synopsis, view ratings and also rate content on their own devices, whether a TV, phone or tablet. However, the emergence of smart TVs, phones and tablets has given rise to “tag while you watch” which enables the user-viewer to tag content and retrieve tagged content using their smart device (Jiang et al., 2013; Perperis and Tsekeridou, 2012). Smart phones and tablets pose constraints on screen size and keyboard and as a result their users usually opt for web applications developed specifically for their device rather than the full web-based system. Smart TVs do not pose such constraints and as a result their users usually opt for add-on applications that are similar to Blu-Ray drop down menus. The plan to institutionalise MC^2 includes participation with any device but this would require additional changes to the MC^2 service to those identified above: first, the viewing subsystem needs to scale up or down to all sizes of smart phone and tablet screens and, secondly, the media management subsystem needs to accommodate live streams and synchronise with the stream in real-time in order to tag time points correctly.

References

AGIUS, H. & ANGELIDES, M. 2007. Closing the content-user gap in MPEG-7: the hanging basket model. Multimedia Systems, 13(2), 155-172.

AGIUS, H. & ANGELIDES, M. 2009. From MPEG-7 user interaction tools to hanging basket models: bridging the gap. *Multimedia Tools and Applications,* 41(3)**,** 375-406.

AGIUS, H. & ANGELIDES, M. C. 2006. MPEG-7 in action: end user experiences with COSMOS-7 front end systems. *In:* the 21st Annual ACM Symposium on Applied Computing (SAC ’06), 23-27 April Dijon, France. New York, NY, USA: ACM Press, 1348-1355.

AHMED, S. I. & GUHA, S. 2012. Distance matters: an exploratory analysis of the linguistic features of Flickr photo tag metadata in relation to impression management. *In:* the 2nd ACM SIGMOD Workshop on Databases and Social Networks, 20 May Scottsdale, Arizona. 2304538: ACM, 7-12.

AL-KHALIFA, H. S. & DAVIS, H. C. 2006. FolksAnnotation: A Semantic Metadata Tool for Annotating Learning Resources Using Folksonomies and Domain Ontologies. *In:* the 3rd International Conference on Innovations in Information Technology, 19-21 November Dubai, UAE. 1-5.

ALLAM, H., BLUSTEIN, J., BLIEMEL, M. & SPITERI, L. 2012. Exploring Factors Impacting Users' Attitude and Intention towards Social Tagging Systems. *In:* the 45th Hawaii International Conference on System Science (HICSS), 4-7 January Grand Wailea, Maui, Hawaii. 3129-3138.

ARNOLD, V., BEDARD, J. C., PHILLIPS, J. R. & SUTTON, S. G. 2012. The impact of tagging qualitative financial information on investor decision making: Implications for XBRL. *International Journal of Accounting Information Systems,* 13(1)**,** 2-20.

AURNHAMMER, M., HANAPPE, P. & STEELS, L. 2006. Augmenting Navigation for Collaborative Tagging with Emergent Semantics *Lecture Notes in Computer Science.* Berlin / Heidelberg: Springer, 58-71.

AYAD, H., ABDULLAH, S. N. H. S. & ABDULLAH, A. 2012. Visual Object Categorization Based on Orientation Descriptor. *In:* Sixth Asia Modelling Symposium (AMS), 29-31 May Bali, Indonesia. 70-74.

BAILER, W., MÉRIALDO, B., KOMPATSIARIS, Y. & CHUA, T.-S. 2012. Semantic multimedia. *Multimedia Tools and Applications,* 59(2)**,** 403-405.

BARRAGÁNS-MARTÍNEZ, A. B., REY-LÓPEZ, M., COSTA MONTENEGRO, E., MIKIC-FONTE, F. A., BURGUILLO, J. C. & PELETEIRO, A. 2010. Exploiting Social Tagging in a Web 2.0

Recommender System. *IEEE Internet Computing,* 14(6), 23-30.

BASTAN, M., CAM, H., GUDUKBAY, U. & ULUSOY, O. 2012. An MPEG-7 Compatible Video Retrieval System with Integrated Support for Complex Multimodal Queries. *IEEE MultiMedia,* PP(99), 1-1.

BATCH, Y., MOHD YUSOF, M., MOHD NOAH, S. A. & LEE, T. P. 2011. MTag: A model to enable collaborative medical tagging in medical blogs. *Procedia Computer Science,* 3(0), 785-790.

BEGELMAN, G., KELLER, P. & SMADJA, F. 2006. Automated Tag Clustering: Improving search and exploration in the tag space. *In:* Workshop on Collaborative Web Tagging, 22 May Edinburgh, UK.

BELLOGÍN, A., CANTADOR, I. & CASTELLS, P. 2013. A comparative study of heterogeneous item recommendations in social systems. *Information Sciences,* 221(0), 142-169.

BENNA, A., MELLAH, H. & HADJARI, K. 2012. Building a social network, based on collaborative tagging, to enhance social information retrieval. *In:* 2012 International Conference on Information Technology and e-Services (ICITeS), 24-26 March Sousse, Tunisia. 1-6.

BLANKINSHIP, E. & MIKHAK, B. 2007. Video-Wikis and Media Fluency. *In:* the 6th International Conference on Interaction Design and Children, 6-8 June Aalborg, Denmark. ACM, 175-176.

BONCHI, F., FRIEDER, O., NARDINI, F. M., SILVESTRI, F. & VAHABI, H. 2012. Interactive and context-aware tag spell check and correction. *In:* the 21st ACM international conference on Information and knowledge management, 29 October - 2 November Maui, Hawaii, USA. 2398534: ACM, 1869-1873.

CANTADOR, I., KONSTAS, I. & JOSE, J. M. 2011. Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web,* 9(1), 1-15.

CARMEL, D., UZIEL, E., GUY, I., MASS, Y. & ROITMAN, H. 2012. Folksonomy-Based Term Extraction for Word Cloud Generation. *ACM Transactions on Intelligent Systems and Technology,* 3(4), 1-20.

CARMINATI, B., FERRARI, E. & PEREGO, A. 2012. A multi-layer framework for personalized social tag-based applications. *Data & Knowledge Engineering,* 79–80(0), 62-86.

CARNEIRO, G. & VASCONCELOS, N. 2005. Formulating semantic image annotation as a supervised learning problem. *In:* the IEEE Computer Society Conference on Computer

Vision and Pattern Recognition, 2005, 20-25 June San Diego, CA. 163-168 vol. 2.

CARVALHO, F. G., TREVISAN, D. G. & RAPOSO, A. 2012. Toward the design of transitional interfaces: an exploratory study on a semi-immersive hybrid user interface. *Virtual Reality,* 16(4)**,** 271-288.

CARVER, J. 2007. The use of grounded theory in empirical software engineering. *In:* the 2006 international conference on Empirical software engineering issues: critical assessment and future directions, Dagstuhl Castle, Germany. 1767419: Springer-Verlag, 42-42.

CELEBI, E. & ALPKOCA, A. 2005. Combining textual and visual clusters for semantic image retrieval and auto-annotation. *In:* the 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, 30 November - 1 December London, UK. 219-225.

CHARMAZ, K. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis,* Newbury Park, CA, Sage Publications Ltd.

CHASTINE, J. W., ZHU, Y. & PRESTON, J. A. 2006. A Framework for Inter-referential Awareness in Collaborative Environments. *In:* International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2006), 17-20 November Atlanta, GA. 1-5.

CHATZILARI, E., NIKOLOPOULOS, S., PATRAS, I. & KOMPATSIARIS, I. 2012. Leveraging social media for scalable object detection. *Pattern Recognition,* 45(8)**,** 2962-2979.

CHOI, K. & SUH, Y. 2013. A new similarity function for selecting neighbors for each target item in collaborative filtering. *Knowledge-Based Systems,* 37(0)**,** 146-153.

CISCO, V. N. I. 2012. *Annual Cisco Visual Networking Index Forecast Projects Global IP Traffic to Increase Fourfold by 2014* [Online]. Available: http://www.cisco.com/en/US/netsol/ns827/networking_solutions_sub_solution.html#~forecast, [Accessed 16/01/2013 2013].

CORBIN, J. & STRAUSS, A. 2008. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory,* Newbury Park, CA, Sage Publications Ltd.

CUI, Y., JIN, J. S., ZHANG, S., LUO, S. & TIAN, Q. 2010. Music video affective understanding using feature importance analysis. *In:* the ACM International Conference on Image and Video Retrieval, 5-7 July Xi'an, China. 1816074: ACM, 213-219.

DALTAYANNI, M., CHUNYE, W. & AKELLA, R. 2012. A Fast Interactive Search System for Healthcare Services. *In:* 2012 Annual SRII Global Conference, 24-27 July San Jose, California, USA. 525-534.

DAS, M., THIRUMURUGANATHAN, S., AMER-YAHIA, S., DAS, G. & YU, C. 2012. Who tags what?: an analysis framework. *Proceedings of the VLDB Endowment,* 5(11), 1567-1578.

DAYLAMANI ZAD, D., ANGELIDES, M. C. & AGIUS, H. 2012. Personalise your Massively Multiplayer Online Game (MMOG) with Artemis. *Multimedia Systems,* 18(1), 69-94.

DAYLAMANI ZAD, D., ANGELIDES, M. C. & AGIUS, H. (Forthcoming). Collaboration through gaming. *In:* ANGELIDES, M. C. & AGIUS, H. (eds.) *Handbook of Digital Games.* Hoboken, New Jersey, USA: Wiley-IEEE Press.

DELLSCHAFT, K. & STAAB, S. 2012. Measuring the influence of tag recommenders on the indexing quality in tagging systems. *In:* the 23rd ACM conference on Hypertext and social media, 25-28 June Milwaukee, Wisconsin, USA. 2310009: ACM, 73-82.

DICKERSON, J. A. & KOSKO, B. 1993. Virtual worlds as fuzzy cognitive maps. *In:* IEEE Virtual Reality Annual International Symposium, 1993., 18-22 Sep 1993 Seattle, WA , USA. 471-477.

DIHUA, X., ZHIJIAN, W., YANLI, Z. & PING, Z. 2012. A Collaborative Tag Recommendation Based on User Profile. *In:* the 4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 26-27 August Nanchang, China. 331-334.

DING, G., WANG, J. & QIN, K. 2010. A visual word weighting scheme based on emerging itemsets for video annotation. *Information Processing Letters,* 110(16), 692-696.

DJUANA, E., XU, Y., LI, Y. & COX, C. 2012. Personalization in tag ontology learning for recommendation making. *In:* the 14th International Conference on Information Integration and Web-based Applications & Services, 3-5 December Bali, Indonesia. 2428804: ACM, 368-377.

DOERFEL, S., JÄSCHKE, R., HOTHO, A. & STUMME, G. 2012. Leveraging publication metadata and social data into FolkRank for scientific publication recommendation. *In:* the 4th ACM RecSys workshop on Recommender systems and the social web, 9-13 September Dublin, Ireland. 2365937: ACM, 9-16.

EBERSMAN, D. A. 2012. Form S-1 Registration Statement: Facebook, Inc. Washington, DC: United States Securities and Exchange Commission.

EDVARDSEN, L. F. H., SØLVBERG, I. T., AALBERG, T. & TRÆTTEBERG, H. 2009. Automatically generating high quality metadata by analyzing the document code of common file types. *In:* the 9th ACM/IEEE-CS Joint Conference on Digital Libraries, 15-19 June Austin, TX, USA. 1555406: ACM, 29-38.

FACH, M. 2011. Google – Under the Hood. *Search Engine Journal* [Online]. Available: http://www.searchenginejournal.com/google-under-the-hood-infographic/37057/.

FEI, C., ANDRITSOS, P., ERKANG, Z. & MILLER, R. J. 2012. AutoDict: Automated Dictionary Discovery. *In:* 2012 IEEE 28th International Conference on Data Engineering (ICDE), 1-5 April Washington, DC, USA. 1277-1280.

FELLBAUM, C. 1998. *WordNet: An Electronic Lexical Database,* Cambridge, MA, MIT Press.

FIGUEIREDO, F., PINTO, H., BELÉM, F., ALMEIDA, J., GONÇALVES, M., FERNANDES, D. & MOURA, E. 2013. Assessing the quality of textual features in social media. *Information Processing & Management,* 49(1)**,** 222-247.

FIRAN, C. S., NEJDL, W. & PAIU, R. 2007. The Benefit of Using Tag-Based Profiles. *In:* 5th Latin American Web Congress (LA-WEB 2007), 31 October – 2 November Santiago, Chile 32-41.

FRIES, J. A., SEGRE, A. M. & POLGREEN, P. M. 2012. Using Online Classified Ads to Identify the Geographic Footprints of Anonymous, Casual Sex-Seeking Individuals. *In:* 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom) Privacy, Security, Risk and Trust (PASSAT), 3-5 September Amsterdam, Netherlands. 402-410.

FU, W.-T., KANNAMPALLIL, T., KANG, R. & HE, J. 2010. Semantic imitation in social tagging. *ACM Transactions on Computer-Human Interaction (TOCHI),* 17(3)**,** 1-37.

FUENTES-LORENZO, D., FERNÁNDEZ, N., FISTEUS, J. A. & SÁNCHEZ, L. 2013. Improving large-scale search engines with semantic annotations. *Expert Systems with Applications,* 40(6)**,** 2287-2296.

GARCÍA, R. D., BENDER, M., ANJORIN, M., RENSING, C. & STEINMETZ, R. 2012. FReSET: an evaluation framework for folksonomy-based recommender systems. *In:* the 4th ACM RecSys workshop on Recommender systems and the social web, 9-13 September Dublin, Ireland. 2365939: ACM, 25-28.

GERBER, A., VAN DER MERWE, A. & ALBERTS, R. 2007. Practical implications of rapid development methodologies. *In:* Computer Science and Information Technology Education Conference, 16-18 November Mauritius. 9.

GLASER, B. G. 1998. *Doing Grounded Theory. Issues and Discussions,* Mill Valley, Ca., Sociology Press, p.9.

GOLDER, S. & HUBERMAN, B. A. 2006a. The Structure of Collaborative Tagging Systems. *HP*

Labs Technical Report.

GOLDER, S. A. & HUBERMAN, B. A. 2006b. Usage patterns of collaborative tagging systems. *Journal of Information Science,* 32(2)**,** 198-208.

GOMBOTZ, R., SCHALL, D., DORN, C. & DUSTDAR, S. 2006. Relevance-Based Context Sharing Through Interaction Patterns. *In:* International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2006), 17-20 November Atlanta, GA 1-7.

GOULDING, C. 1998. Grounded theory: the missing methodology on the interpretivist agenda. *Qualitative Market Research,* 1(1)**,** 50 - 57, p.51.

GREEN, N., BREIMYER, P., KUMAR, V. & SAMATOVA, N. F. 2010. PackPlay: mining semantic data in collaborative games. *In:* the Fourth Linguistic Annotation Workshop, 15-16 July Uppsala, Sweden. 1868757: Association for Computational Linguistics, 227-234.

GRIMM, V. & MENGEL, F. 2012. An experiment on learning in a multiple games environment. *Journal of Economic Theory,* 147(6)**,** 2220-2259.

GROSSMAN, L. 2012. The Beast with a Billion Eyes. *TIME.*

HAN, S., HAN, J.-J., KIM, J. D. K. & KIM, C. 2013. Connecting users to virtual worlds within MPEG-V standardization. *Signal Processing: Image Communication,* 28(2)**,** 97-113.

HAQ, M. U., AHMED, H. & QAMAR, A. M. 2012. Dynamic entity and relationship extraction from news articles. *In:* 2012 International Conference on Emerging Technologies (ICET), 8-9 October Islamabad, Pakistan. 1-5.

HASLHOFER, B. & KLAS, W. 2010. A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys (CSUR),* 42(2)**,** 1-37.

HONGBO, S., WENHUI, F., WEIMING, S. & TIANYUAN, X. 2013. Ontology Fusion in High-Level-Architecture-Based Collaborative Engineering Environments. *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* 43(1)**,** 2-13.

HSIEH, W.-T., STU, J., CHEN, Y.-L. & CHOU, S.-C. T. 2009. A collaborative desktop tagging system for group knowledge management based on concept space. *Expert Systems with Applications,* 36(5)**,** 9513-9523.

HSU, Y.-C., JENG, T.-S., SHEN, Y.-T. & CHEN, P.-C. 2012. SynTag: a web-based platform for labeling real-time video. *In:* the ACM 2012 conference on Computer Supported Cooperative Work, 11-15 February Seattle, Washington, USA. 2145312: ACM, 715-718.

HU, J. & LAM, K.-M. 2013. An efficient two-stage framework for image annotation. *Pattern Recognition,* 46(3)**,** 936-947.

HUCK, S., JEHIEL, P. & RUTTER, T. 2011. Feedback spillover and analogy-based expectations: A multi-game experiment. *Games and Economic Behavior,* 71(2)**,** 351-365.

HYUN-SEOK, M., DE NEVE, W. & YONG MAN, R. 2010. Towards using semantic features for near-duplicate video detection. *In:* the 2010 IEEE International Conference on Multimedia and Expo, 19-23 July Suntec City, Singapore. 1364-1369.

HYUN-SEOK, M., JAE YOUNG, C., DE NEVE, W. & YONG MAN, R. 2012. Near-Duplicate Video Clip Detection Using Model-Free Semantic Concept Detection and Adaptive Semantic Distance Measurement. *IEEE Transactions on Circuits and Systems for Video Technology,* 22(8)**,** 1174-1187.

IOAKIM, G., KYRIACOU, E., SOFOKLEOUS, A. A., CHISTODOULOU, C. & PATTICHIS, C. S. 2012. An MPEG-7 image retrieval system of atherosclerotic carotid plaque images. *In:* 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), 11-13 November Larnaca, Cyprus. 512-516.

ISO/IEC 2003. Information Technology – Multimedia Content Description Interface – Part 5: Multimedia Description Schemes. Geneva, Switzerland.

ISO/IEC 2004. Information Technology – Multimedia Content Description Interface – Part 5: Multimedia Description Schemes: Amendment 2: Multimedia Description Schemes Extensions. Geneva, Switzerland.

ISO/IEC 2005a. Information Technology – Multimedia Content Description Interface – Part 5: Multimedia Description Schemes: Amendment 2: Multimedia Description Schemes User Preference Extensions. Geneva, Switzerland.

ISO/IEC 2005b. Information Technology – Multimedia Content Description Interface – Part 9: Profiles and Levels. Geneva, Switzerland.

ISOTANI, S., MIZOGUCHI, R., ISOTANI, S., CAPELI, O. M., ISOTANI, N., DE ALBUQUERQUE, A. R. P. L., BITTENCOURT, I. I. & JAQUES, P. 2013. A Semantic Web-based authoring tool to facilitate the planning of collaborative learning scenarios compliant with learning theories. *Computers & Education,* 63(0)**,** 267-284.

JACKSON, T. W. & SMITH, S. 2012. Retrieving relevant information: traditional file systems versus tagging. *Journal of Enterprise Information Management,* 25(1)**,** 79-93.

JAIN, A. K., MURTY, M. N. & FLYNN, P. J. 1999. Data clustering: a review. *ACM Comput. Surv.,* 31(3)**,** 264-323.

JIAN, J. & QUN, C. 2012. A trust-based Top-K recommender system using social tagging network. *In:* the 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 29-31 May Chongqing, China. 1270-1274.

JIANG, T., LI, J., WU, R. & XIANG, K. 2013. TVEar: A TV-tagging System Based on Audio Fingerprint. *In:* LI, S., SADDIK, A., WANG, M., MEI, T., SEBE, N., YAN, S., HONG, R. & GURRIN, C. (eds.) *Advances in Multimedia Modeling.* Springer Berlin Heidelberg, 528-531.

JOHNSON, J. R., MILLER, A., KHAN, L. & THURAISINGHAM, B. 2012. Extracting semantic information structures from free text law enforcement data. *In:* 2012 IEEE International Conference on Intelligence and Security Informatics (ISI), 11-14 June Washington, DC, USA. 177-179.

JUNG, Y. 2008. Influence of Sense of Presence on Intention to Participate in a Virtual Community. *In:* the 41st Annual Hawaii International Conference on System Sciences, 7-10 January Waikoloa, Big Island, Hawaii. IEEE Computer Society.

KELLE, U. 2005. Emergence vs. Forcing of empirical data? A Crucial Problem of Grounded Theory Reconsidered. *Qualitative Social Research,* 6(2).

KERMARREC, A.-M., ROWSTRON, A., SHAPIRO, M. & DRUSCHEL, P. 2001. The IceCube approach to the reconciliation of divergent replicas. *In:* Twentieth Annual ACM Symposium on Principles of Distributed Computing, 26-29 August Newport, Rhode Island, United States. ACM.

KIM, H.-N., ALKHALDI, A., EL SADDIK, A. & JO, G.-S. 2011. Collaborative user modeling with user-generated tags for social recommender systems. *Expert Systems with Applications,* 38(7)**,** 8488-8496.

KOSKO, B. 1986. Fuzzy Cognitive Maps. *International Journal of Man-Machine Studies,* 24(56-74.

KOSKO, B. 1997. *Fuzzy engineering,* Upper Saddle River, NJ, USA, Prentice-Hall, Inc.

KUMAR, R. & TOMKINS, A. 2010. A characterization of online browsing behavior. *In:* the 19th International Conference on World Wide Web, 26-30 April Raleigh, North Carolina, USA. 1772748: ACM, 561-570.

KWON, O. 2009. A two-step approach to building bilateral consensus between agents based on relationship learning theory. *Expert Systems with Applications,* 36(9)**,** 11957-11965.

LABORIE, S., MANZAT, A.-M. & SÉDES, F. 2010. A generic framework for the integration of heterogeneous metadata standards into a multimedia information retrieval system.

In: the Adaptivity, Personalization and Fusion of Heterogeneous Information, 28-30 April Paris, France. 1937073: Le Centre de Hautes Etudes Internationales d'Informatique Documentaire, 80-83.

LAU, S. B. Y. & CHIEN-SING, L. 2012. Enhancing Collaborative Filtering of Learning Resources with Semantically-Enhanced Social Tags. *In:* 2012 IEEE 12th International Conference on Advanced Learning Technologies (ICALT), 4-6 July Rome, Italy. 281-285.

LEE, S.-S. & YONG, H.-S. 2007. TagPlus: A Retrieval System using Synonym Tag in Folksonomy. *In:* International Conference on Multimedia and Ubiquitous Engineering (MUE '07), 26-28 April Seoul, Korea 294-298.

LEE, S., NEVE, W. D. & RO, Y. M. 2012. Towards data-driven estimation of image tag relevance using visually similar and dissimilar folksonomy images. *In:* the 2012 international workshop on Socially-aware multimedia, 29 October - 2 November Nara, Japan. 2390880: ACM, 3-8.

LEI, L. 2012. A novel violent videos classification scheme based on the bag of audio words features [Document Suppressed in IEEE Xplore]. *In:* Ninth International Conference on Information Technology: New Generations (ITNG), 16-18 April Las Vegas, Nevada, USA. 7-13.

LI, Q. & LU, S. C. Y. 2008. Collaborative Tagging Applications and Approaches. *IEEE Multimedia,* 15(3)**,** 14-21.

LI, Z., SHI, Z., LIU, X. & SHI, Z. 2011. Modeling continuous visual features for semantic image annotation and retrieval. *Pattern Recognition Letters,* 32(3)**,** 516-523.

LIPCZAK, M., SIGURBJORNSSON, B. & JAIMES, A. 2012. Understanding and leveraging tag-based relations in on-line social networks. *In:* the 23rd ACM conference on Hypertext and social media, 25-28 June Milwaukee, Wisconsin, USA. 2310035: ACM, 229-238.

LU, Y. & LI, Z.-N. 2008. Automatic object extraction and reconstruction in active video. *Pattern Recogn.,* 41(3)**,** 1159-1172.

MALEEWONG, K., ANUTARIYA, C. & WUWONGSE, V. 2008. A Collective Intelligence Approach to Collaborative Knowledge Creation. *In:* the 4th International Conference on Semantics, Knowledge and Grid, 3-5 December Beijing, China. IEEE Computer Society, 64-70.

MANIU, S. & CAUTIS, B. 2012. Taagle: efficient, personalized search in collaborative tagging networks. *In:* the 2012 ACM SIGMOD International Conference on Management of Data, 20-24 May Scottsdale, Arizona, USA. 2213926: ACM, 661-664.

MATAVIRE, R. & BROWN, I. 2008. Investigating the use of "Grounded Theory" in information systems research. *In:* Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT research in Developing Countries: Riding the Wave of Technology, 6-8 October Wilderness, South Africa. ACM, 139-147.

MAURER, F. & MARTEL, S. 2002. Extreme programming. Rapid development for Web-based applications. *IEEE Internet Computing,* 6(1)**,** 86-90.

MAYERNIK, M. S., BATCHELLER, A. L. & BORGMAN, C. L. 2011. How institutional factors influence the creation of scientific metadata. *In:* the 2011 iConference, 8-11 February Seattle, Washington. 1940818: ACM, 417-425.

MCNALLY, K., O, M. P., MAHONY, COYLE, M., BRIGGS, P. & SMYTH, B. 2011. A Case Study of Collaboration and Reputation in Social Web Search. *ACM Transactions on Intelligent Systems and Technology (TIST),* 3(1)**,** 1-29.

MEZGHANI, M., ZAYANI, C. A., AMOUS, I. & GARGOURI, F. 2012. A user profile modelling using social annotations: a survey. *In:* the 21st international conference companion on World Wide Web, 16-20 April Lyon, France. 2188230: ACM, 969-976.

MICROSOFT. 2012a. *Expression Encoder 4 Pro: Overview* [Online]. Available: http://www.microsoft.com/expression/products/EncoderPro_Overview.aspx.

MICROSOFT. 2012b. What is Silverlight. Available: http://www.microsoft.com/silverlight/what-is-silverlight/.

MIKA, P. 2007. Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web,* 5(1)**,** 5-15.

MISHRA, S., GORAI, A., OBEROI, T. & GHOSH, H. 2010. Efficient Visualization of Content and Contextual Information of an Online Multimedia Digital Library for Effective Browsing. *In:* the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, August 31 - September 3 Toronto, Ontario, Canada 257-260.

MONEY, A. G. & AGIUS, H. 2008. Video summarisation: A conceptual framework and survey of the state of the art. *Journal of Visual Communication and Image Representation,* 19(2)**,** 121-143.

MSDN. 2010. LINQ (Language-Integrated Query). Available: http://msdn.microsoft.com/en-us/library/bb397926.aspx.

MYUNG JIN, C., TORRALBA, A. & WILLSKY, A. S. 2012. A Tree-Based Context Model for Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 34(2)**,** 240-252.

NAAMAN, M. & NAIR, R. 2008. ZoneTag's Collaborative Tag Suggestions: What is This Person Doing in My Phone? *IEEE Multimedia,* 15(3)**,** 34-40.

NÉVÉOL, A., ISLAMAJ DOGAN, R. & LU, Z. 2011. Semi-automatic semantic annotation of PubMed queries: A study on quality, efficiency, satisfaction. *Journal of Biomedical Informatics,* 44(2)**,** 310-318.

NING, Z., CHEUNG, W. K., GUOPING, Q. & XIANGYANG, X. 2011. A Hybrid Probabilistic Model for Unified Collaborative and Content-Based Image Tagging. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 33(7)**,** 1281-1294.

OBJECT MANAGEMENT GROUP 2011. OMG Unified Modeling Language (OMG UML), Superstructure. Version 2.4.1 ed.

OHMUKAI, I., HAMASAKI, M. & TAKEDA, H. 2005. A Proposal of Community-Based Folksonomy with RDF Metadata. *In:* the 4th International Semantic Web Conference (ISWC), 6-10 November Galaway, Ireland.

OJOKOH, B., ZHANG, M. & TANG, J. 2011. A trigram hidden Markov model for metadata extraction from heterogeneous references. *Information Sciences,* 181(9)**,** 1538-1551.

OTSUKA, I., NAKANE, K., DIVAKARAN, A., HATANAKA, K. & OGAWA, M. 2005. A highlight scene detection and video summarization system using audio feature for a personal video recorder. *IEEE Transactions on Consumer Electronics,* 51(1)**,** 112-116.

PAN, W., CHEN, S. & FENG, Z. 2012. Investigating the Collaborative Intention and Semantic Structure among Co-occurring Tags Using Graph Theory. *In:* 2012 IEEE 16th International Enterprise Distributed Object Computing Conference Workshops (EDOCW), 10-14 September Beijing, China. 190-195.

PAPADOPOULOU, S., IGNAT, C., OSTER, G. & NORRIE, M. 2006. Increasing Awareness in Collaborative Authoring through Edit Profiling. *In:* International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2006), 17-20 November Atlanta, GA 1-9.

PARRA-ARNAU, J., PEREGO, A., FERRARI, E., FORNÉ, J. & REBOLLO-MONEDERO, D. 2013. Privacy-Preserving Enhanced Collaborative Tagging. *IEEE Transactions on Knowledge and Data Engineering,* PP(99)**,** 1-1.

PARRA-ARNAU, J., REBOLLO-MONEDERO, D., FORNÉ, J., MUÑOZ, J. L. & ESPARZA, O. 2012.

Optimal tag suppression for privacy protection in the semantic Web. *Data & Knowledge Engineering,* 81–82(0), 46-66.

PASCHALAKIS, S., IWAMOTO, K., BRASNETT, P., SPRLJAN, N., OAMI, R., NOMURA, T., YAMADA, A. & BOBER, M. 2012. The MPEG-7 Video Signature Tools for Content Identification. *IEEE Transactions on Circuits and Systems for Video Technology,* 22(7), 1050-1063.

PEA, R., LINDGREN, R. & ROSEN, J. 2006. Computer-supported collaborative video analysis. *In:* the 7th International Conference on Learning Sciences, 27 June - 1 July Bloomington, Indiana. International Society of the Learning Sciences, 516-521.

PERPERIS, T. & TSEKERIDOU, S. 2012. TV Content Analysis and Annotation for Parental Control. *In:* LIAN, S. (ed.) *TV Content Analysis.* Boca Raton, FL, USA: Auerbach Publications.

PRINCETON UNIVERSITY. 2010. *About WordNet* [Online]. Princeton University. Available: http://wordnet.princeton.edu.

QIN, C., BAO, X., CHOUDHURY, R. & NELAKUDITI, S. 2012. TagSense: Leveraging Smartphones for Automatic Image Tagging. *IEEE Transactions on Mobile Computing,* Early Access(

QIU, J., SU, J., LUO, L. & ZHOU, C. 2008. The Application of Multimedia in Academic Library. *In:* the 2008 International Conference on MultiMedia and Information Technology, 30-31 December Three Gorges, China. 703-706.

RAZIKIN, K., GOH, D. H., CHUA, A. Y. K. & LEE, C. S. 2011. Social tags for resource discovery: a comparison between machine learning and user-centric approaches. *Journal of Information Science,* 37(4), 391-404.

RODENHAUSEN, T., ANJORIN, M., GARCÍA, R. D. & RENSING, C. 2012. Context determines content: an approach to resource recommendation in folksonomies. *In:* the 4th ACM RecSys workshop on Recommender systems and the social web, 9-13 September Dublin, Ireland. 2365938: ACM, 17-24.

RODRIGUEZ, M. A., BOLLEN, J. & SOMPEL, H. V. D. 2009. Automatic metadata generation using associative networks. *ACM Transactions on Information Systems (TOIS),* 27(2), 1-20.

RYU, J., SOHN, Y. & KIM, M. 2002. MPEG-7 metadata authoring tool. *In:* the 10th ACM international conference on Multimedia, 1-6 December Juan-les-Pins, France. 641061: ACM, 267-270.

SAATHOFF, C. & SCHERP, A. 2010. Unlocking the semantics of multimedia presentations in the web with the multimedia metadata ontology. *In:* the 19th International Conference on World Wide Web, 26-30 April Raleigh, North Carolina, USA. 1772775: ACM, 831-840.

SAN PEDRO, J., SIERSDORFER, S. & SANDERSON, M. 2011. Content redundancy in YouTube and its application to video tagging. *ACM Transactions on Information Systems (TOIS),* 29(3)**,** 1-31.

SANTOS, F. R., DA COSTA CORDEIRO, W. L., GASPARY, L. P. & BARCELLOS, M. P. 2013. Beyond pollution and taste: A tag-based strategy to increase download quality in P2P file sharing systems. *Computer Communications,* 36(2)**,** 191-202.

SARMA, A., NOROOZI, Z. & HOEK, A. V. D. 2003. Palantir: raising awareness among configuration management workspaces. *In:* 25th International Conference on Software Engineering, 3-10 May Portland, Oregon. IEEE Computer Society, 444-454.

SCHOEFEGGER, K. & GRANITZER, M. 2012. Overview and analysis of personal and social tagging context to construct user models. *In:* the 2nd Workshop on Context-awareness in Retrieval and Recommendation, 14 February Lisbon, Portugal. 2162106: ACM, 14-21.

SCHÖFEGGER, K., KÖRNER, C., SINGER, P. & GRANITZER, M. 2012. Learning user characteristics from social tagging behavior. *In:* the 23rd ACM conference on Hypertext and social media, 25-28 June Milwaukee, Wisconsin, USA. 2310031: ACM, 207-212.

SEMERARO, G., LOPS, P., GEMMIS, M. D., MUSTO, C. & NARDUCCI, F. 2012. A folksonomy-based recommender system for personalized access to digital artworks. *Journal on Computing and Cultural Heritage,* 5(3)**,** 1-22.

SEO, D. W. & LEE, J. Y. 2013. Physical query interface for tangible augmented tagging and interaction. *Expert Systems with Applications,* 40(6)**,** 2032-2042.

SHEN, H. & SUN, C. 2005. Syntax-based reconciliation for asynchronous collaborative writing. *In:* International Conference on Collaborative Computing: Networking, Applications and Worksharing, 19-21 December San Jose, CA. 10.

SHIN, Y., KIM, Y. & KIM, E. Y. 2010. Automatic textile image annotation by predicting emotional concepts from visual features. *Image and Vision Computing,* 28(3)**,** 526-537.

SHIRMOHAMMADI, S. & CLAYPOOL, M. 2009. Massively multiplayer online gaming systems and applications. *Multimedia Tools and Applications,* 45(1-3)**,** 1-5.

SHOTTON, J., SHARP, T., KIPMAN, A., FITZGIBBON, A., FINOCCHIO, M., BLAKE, A., COOK, M. & MOORE, R. 2013. Real-time human pose recognition in parts from single depth images. *Communications of the ACM,* 56(1)**,** 116-124.

SINGH, C., SARKAR, S., ARAM, A. & KUMAR, A. 2012. Cooperative profit sharing in coalition-based resource allocation in wireless networks. *IEEE/ACM Transactions on Networking (TON),* 20(1)**,** 69-83.

SMOLIAR, S. W. & HONGJIANG, Z. 1994. Content based video indexing and retrieval. *IEEE Multimedia,* 1(2)**,** 62-72.

SOLSKINNSBAKK, G., GULLA, J. A., HADERLEIN, V., MYRSETH, P. & CERRATO, O. 2012. Quality of hierarchies in ontologies and folksonomies. *Data & Knowledge Engineering,* 74(0)**,** 13-25.

STROHMAIER, M., KÖRNER, C. & KERN, R. 2012. Understanding why users tag: A survey of tagging motivation literature and results from an empirical study. *Web Semantics: Science, Services and Agents on the World Wide Web,* 17(0)**,** 1-11.

TAN, S. & LUKOSE, D. 2012. Cognitive Semantic Model for Visual Object Recognition in Image. *In:* WANG, F., LEI, J., LAU, R. H. & ZHANG, J. (eds.) *Multimedia and Signal Processing.* Springer Berlin Heidelberg, 67-78.

TANG, L., LIU, H., ZHANG, J., AGARWAL, N. & SALERNO, J. J. 2008. Topic taxonomy adaptation for group profiling. *ACM Trans. Knowl. Discov. Data,* 1(4)**,** 1-28.

TANG, L., WANG, X. & LIU, H. 2011. Group Profiling for Understanding Social Structures. *ACM Transactions on Intelligent Systems and Technology (TIST),* 3(1)**,** 1-25.

TAO, C. & EMBLEY, D. W. 2009. Automatic hidden-web table interpretation, conceptualization, and semantic annotation. *Data & Knowledge Engineering,* 68(7)**,** 683-703.

THORNTON, K. & MCDONALD, D. W. 2012. Tagging Wikipedia: collaboratively creating a category system. *In:* the 17th ACM international conference on Supporting group work, 27-31 October Sanibel Island, Florida, USA. 2389210: ACM, 219-228.

TIAN, Y. 2006. *Developing an open source software development process model using grounded theory.* University of Nebraska at Lincoln.

TONELLI, S., GIULIANO, C. & TYMOSHENKO, K. 2013. Wikipedia-based WSD for multilingual frame annotation. *Artificial Intelligence,* 194(0)**,** 203-221.

TOTI, D., ATZENI, P. & POLTICELLI, F. 2012. A Knowledge Discovery Methodology for Semantic Categorization of Unstructured Textual Sources. *In:* 2012 Eighth

International Conference on Signal Image Technology and Internet Based Systems (SITIS), 25-29 November Sorrento - Naples, Italy. 944-951.

TOURNÉ, N. & GODOY, D. 2012. Evaluating tag filtering techniques for web resource classification in folksonomies. *Expert Systems with Applications,* 39(10)**,** 9723-9729.

TRONCY, R., BAILER, W., HÖFFERNIG, M. & HAUSENBLAS, M. 2010. VAMP: A Service for Validating MPEG-7 Descriptions w.r.t. to Formal Profile Definitions. *Multimedia Tools and Applications,* 46(2-3)**,** 307-329.

TSAI, D., JING, Y., ROWLEY, H., LIU, Y., IOFFE, S. & REHG, J. 2011. Large-scale image annotation using visual synset. *In:* IEEE International Conference on Computer Vision (ICCV 2011), 6-13 November Barcelona, Spain. 611-618.

TSINGALIS, I., VRETOS, N., NIKOLAIDIS, N. & PITAS, I. 2012. Anthropocentric descriptors and description schemes for multi-view video content. *In:* 16th IEEE Mediterranean Electrotechnical Conference (MELECON), 25-28 March Medina Yasmine Hammamet, Tunisia 133-136.

UDDIN, M. N., DUONG, T. H., NGUYEN, N. T., QI, X.-M. & JO, G. S. 2013. Semantic similarity measures for enhancing information retrieval in folksonomies. *Expert Systems with Applications,* 40(5)**,** 1645-1653.

ULGES, A., SCHULZE, C., KEYSERS, D. & BREUEL, T. 2008. A System That Learns to Tag Videos by Watching Youtube. *Computer Vision Systems.* 415-424.

VANDIC, D., VAN DAM, J.-W. & FRASINCAR, F. 2012. A semantic-based approach for searching and browsing tag spaces. *Decision Support Systems,* 54(1)**,** 644-654.

VIITANIEMI, V. & LAAKSONEN, J. 2007. Evaluating the performance in automatic image annotation: Example case by adaptive fusion of global image features. *Signal Processing: Image Communication,* 22(6)**,** 557-568.

VIJAY, V. & JACOB, I. J. 2012. Combined approach of user specified tags and content-based image annotation. *In:* 2012 International Conference on Devices, Circuits and Systems (ICDCS), 15-16 March Coimbatore, Tamil Nadu, India. 162-166.

VOSS, J. 2007. Tagging, Folksonomy & Co - Renaissance of Manual Indexing? *In:* the 10th International Symposium for Information Science, 30 May - 1 June Cologne, Germany. 234–254.

VRETOS, N., NIKOLAIDIS, N. & PITAS, I. 2012. The use of Audio-Visual Description Profile in 3D video content description. *In:* 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 15-17 October ETH Zurich, Switzerland. 1-4.

WALTL, M., RAINER, B., TIMMERER, C. & HELLWAGNER, H. 2013. An end-to-end tool chain for Sensory Experience based on MPEG-V. *Signal Processing: Image Communication,* 28(2)**,** 136-150.

WILLIAM, K. 2006. Exploiting "The World is Flat" Syndrome in Digital Photo Collections for Contextual Metadata. *In:* the 8th IEEE International Symposium on Multimedia (ISM'06), 11-13 December San Diego, CA 341-347.

WILLIAMS, L., MENEELY, A. & SHIPLEY, G. 2010. Protection Poker: The New Software Security "Game". *IEEE Security & Privacy,* 8(3)**,** 14-20.

WU, R.-S. & LI, P.-C. 2011. Video annotation using hierarchical Dirichlet process mixture model. *Expert Systems with Applications,* 38(4)**,** 3040-3048.

YAMAMOTO, D., MASUDA, T., OHIRA, S. & NAGAO, K. 2008. Video Scene Annotation Based on Web Social Activities. *IEEE Multimedia,* 15(3)**,** 22-32.

YANG, D., WU, D., KOOLMANOJWONG, S., BROWN, A. W. & BOEHM, B. W. 2008. WikiWinWin: A Wiki Based System for Collaborative Requirements Negotiation. *In:* the 41st Annual Hawaii International Conference on System Sciences, 7-10 January Waikoloa, Big Island, Hawaii. 24-24.

YANG, Y.-H., BOGDANOV, D., HERRERA, P. & SORDO, M. 2012. Music retagging using label propagation and robust principal component analysis. *In:* 21st international conference companion on World Wide Web, 16-20 April Lyon, France. 2188217: ACM, 869-876.

YONG-HWAN, L., SANG-BURM, R. & BONAM, K. 2012. Content-Based Image Retrieval Using Wavelet Spatial-Color and Gabor Normalized Texture in Multi-resolution Database. *In:* Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 4-6 July Palermo, Italy. 371-377.

YOON, K. 2013. End-to-end framework for 4-D broadcasting based on MPEG-V standard. *Signal Processing: Image Communication,* 28(2)**,** 127-135.

YOU, J., LIU, G. & PERKIS, A. 2010. A semantic framework for video genre classification and event analysis. *Signal Processing: Image Communication,* 25(4)**,** 287-302.

YOUTUBE. 2012. Press Room: Statistics. Available: http://www.youtube.com/t/press_statistics.

YU, L., YI, C., GUANG, Y., HONG-KE, Z., JUN-TING, C. & HUA-QING, M. 2012. Mining latent tag group based on tag dependency relation for recommendation in collaborative tagging systems. *In:* 2012 International Conference on Machine Learning and Cybernetics (ICMLC), 15-17 July Xian, China. 102-106.

YUILL, N. & ROGERS, Y. 2012. Mechanisms for collaboration: A design and evaluation framework for multi-user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI),* 19(1)**,** 1-25.

ZHANG, X., XU, C., CHENG, J., LU, H. & MA, S. 2009. Effective Annotation and Search for Video Blogs with Integration of Context and Content Analysis. *IEEE Transactions on Multimedia,* 11(2)**,** 272-285.

ZHANG, X., ZHAO, X., LI, Z., XIA, J., JAIN, R. & CHAO, W. 2013. Social image tagging using graph-based reinforcement on multi-type interrelated objects. *Signal Processing,* In Press, Corrected Proof(

ZHANG, Y., ZHANG, B., GAO, K., GUO, P. & SUN, D. 2012. Combining content and relation analysis for recommendation in social tagging systems. *Physica A: Statistical Mechanics and its Applications,* 391(22)**,** 5759-5768.

ZHENG, Y.-T., ZHAO, M., NEO, S.-Y., CHUA, T.-S. & TIAN, Q. 2008. Visual Synset: Towards a Higher-level Visual Representation. *In:* IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), 23-28 June Anchorage, Alaska, USA. 1-8.

ZHOU, H., SADKA, A. H., SWASH, M. R., AZIZI, J. & SADIQ, U. A. 2010. Feature extraction and clustering for dynamic video summarisation. *Neurocomputing,* 73(10-12)**,** 1718-1729.

ZHU, Y.-T., LIU, S.-H., CHENG, X.-Q., LIU, Y., WANG, Y.-Z. & LIU, J.-G. 2012. Using wordnet-based neighborhood for improving social tag recommendation. *In:* the 8th international conference on Intelligent Computing Theories and Applications, 25-29 July Huangshan, China. 2364227: Springer-Verlag, 221-228.

Appendix I: Initial Experiment

Tasks

Each participant was required to tag and content model:

- Approximately 15 images from Flickr.
- Approximately 10 web pages containing AV content from del.icio.us.
- 3-5 minutes worth of videos (either one 5 minute long video or a number of videos that total 5 minutes together) from YouTube and COSMOSIS.

This ensured that users need not take more than about an hour to complete the tasks, since more time than this would have greatly discouraged them from participating initially or from completing all tasks. At the same time, the video duration and number of images and web pages were sufficient to accommodate meaningful semantics. Users did not have to complete all the tasks in one session. Each user had two weeks to complete the tasks. Interviews commenced as soon as the first user finished their tasks. For YouTube, Flickr and del.icio.us, users added an MC^2 user account as their friend. This enabled post-experiment collection of tags. The COSMOSIS system outputs a file containing the metadata that was sent by the user via email to the investigator. It was stressed to users that communication with each other is vital for the experiment and utility of the results would be greatly reduced otherwise.

Image tasks

In order to accommodate a range of common situations so that comparisons and contrasts between them may be drawn, users were presented with three batches of images that share similar content features and were all based around the same concept or theme.

- **Image Batch 1:** Users did not have any guidelines and could tag freely.
- **Image Batch 2:** Users were told to tag images based on a set of features that are exhibited (from the feature list detailed in Section 2.2).

- **Image Batch 3:** This batch was based on an alternative set of features and was used to uncover differences in tagging from the previous two batches. Users were asked to describe a set of 5 photos which they thought were similar to the previous two batches.

Each batch contained 5 images and thus did not take more than 10 minutes, which meant the whole image tagging task took between 15-30 minutes. Since each participant needed 15 images, 750 images were needed in total. However, not all participants used the images provided. Therefore, approximately 300 images were deemed sufficient.

Web pages with AV content

Users were asked to tag in del.icio.us 10 websites that feature AV content. They were asked to find websites matching their own interests and the group they were assigned to, based on their own searches or previous knowledge. The queries were recorded by the users for later analysis since the keywords may reveal how the user *expected* the AV content to be tagged and may also reveal their priorities in describing AV content. Users were allowed around 2 minutes per website, therefore, this set of tasks took approximately 15-20 minutes in total.

Videos

Users were asked to upload and tag or just tag in YouTube 3-5 minutes worth of video. This may have either been a single video of 5 minutes duration or multiple videos adding up to 5 minutes. Users were then asked to content model the same video(s) in COSMOSIS. A length of 5 minutes is sufficient to accommodate a richness of semantics while also not overstretching the user. It was estimated that tagging the 3-5 minute long video would take about 1-2 minutes in YouTube, while content modelling in COSMOSIS would take about 6-10 minutes.

User groups

The section presents the tasks undertaken by each user group in detail. In all four groups, websites were chosen by the user.

Group 1: Personal

This group was asked to use their own photos and videos. The following features are most relevant to this group and thus content was selected which is rich in these features (although other features are likely to also be exhibited):

- **Objects:** People, Animals, Inanimate Objects, Properties
- **Events:** Including Music and Noise
- **Relationships:** Spatial, Location

Group 2: Business

This group was provided with business-oriented AV material. The following features are most relevant to this group and thus content has been selected which is rich in these features (although other features are likely to also be exhibited):

- **Objects:** People, Animals, Inanimate Objects, Properties
- **Events:** Including Conversation
- **Relationships:** User, Part, Specializes

Group 3: Academic

This group was provided with AV content of an academic nature. The following features are most relevant to this group and thus content has been selected which is rich in these features (although other features are likely to also be exhibited):

- **Objects:** People, Animals, Inanimate Objects, Properties
- **Events:** Including Noise
- **Relationships:** Temporal, Spatial, Causer, User, Part, Specializes

Group 4: Recreational

This group was provided with a set of recreational AV content. The following features are most relevant to this group and thus content has been selected which is rich in these features (although other features are likely to also be exhibited):

- **Objects:** People, Animals, Inanimate Objects, Properties
- **Events:** Including Music, Noise, Conversation
- **Relationships:** Causer, Part, Location

Table I.1 illustrates the relationships between the content categories, user groups and content features.

Table I.1. Mapping of content categories to user groups to content features.

	Personal User Group 1	Business User Group 2	Academic User Group 3	Recreation User Group 4
People	X	X	X	X
Inanimate Objects	X	X	X	X
Animals	X	X	X	X
Properties	X	X	X	X
Events	X	X	X	X
Music*	X			X
Noise*	X		X	X
Conversation*		X		X
Temporal Relations*			X	
Spatial Relations	X		X	
Causer Relations			X	X
User Relations		X	X	
Part Relations		X	X	X
Specializes Relations		X	X	
Location Relations	X			X

** Applies only to video, not images.*

Appendix II: Experiment Consent Form

The experiment undertaken is part of the EPSRC-funded MC^2 research project at Brunel University. Participation in this experiment involves undertaking a number of tasks where audio-visual files are described using a custom software system (which will be provided) and the following websites: YouTube, Flickr and del.icio.us. We will collect the descriptions that participants create while using the software and the websites. The tasks will take approximately an hour to complete in total and do not have to be completed in a single session. Participants will communicate electronically in small groups while undertaking the tasks and this communication will be recorded. At a convenient time after completion of the tasks, participants will be individually interviewed briefly and encouraged to discuss their experiences.

All information that you provide to us will be in confidence. However, you will have an opportunity to review all data, including the communication transcripts, before they are made available to us.

Please note that your participation in this study is completely voluntary and granting your consent does not obligate you to complete the experiment. You are completely free to withdraw at any time if you are unable or unwilling to continue.

If you do not fully understand any part of the experiment or require further information, please do not hesitate to ask us.

In accordance with the guidelines provided by Brunel University's Research Ethics Committee, we are required to obtain a record of your consent to take part in this experiment. By signing and dating below, you confirm that you have been informed of the nature of the research and your rights to withdraw and that you agree to take part in the experiment.

Thank you for giving your time to MC^2. We appreciate it.

Appendix III: Repeat Experiment

Each participant was required to tag and content model using the MC^2 online service:

- 6-10 minutes worth of videos (two separate videos, each of 3-5 minutes duration)
- One audio file of approximately 3-5 minutes duration

This was to ensure that users need not take more than about an hour to complete the tasks, since more time than this would greatly discourage them from participating, either initially or in completing all tasks. At the same time, the durations were sufficient to accommodate meaningful semantics. Users did not have to complete all the tasks in one session. Each user had six weeks to complete the tasks. Interviews started as soon as the first user finished their tasks. The MC^2 online service stored all the transactions of the user with the system and the researcher referred to this logging system in order to determine the results and data.

New media

Users were asked to upload and tag or just tag in MC^2 online service, either two videos of 3-5 minutes duration (together 6-10 minutes) or one video totalling 6-10 minutes. Group 1 was asked to choose one video and one audio file from their own personal files of around 3-5 minutes duration. This helped demonstrate the users' normal behaviour towards modelling this type of content which is inherently personal to them. They were then asked to choose any video of 3-5 minutes duration but not necessarily from any particular content category. This ensured that they chose a video that they were comfortable modelling, therefore increasing the interaction with the service and also providing the researcher with a view of what type of videos users were more comfortable modelling.

Groups 2-4 used a video of 3-5 minutes duration and an audio file of 3-5 minutes within their relevant content category both of which were provided by the researcher. Then, similar to Group 1, they were asked to choose a video on any topic or content that they were comfortable with and model it using the service. This video should have been of 3-5 minutes duration. These durations were sufficient to accommodate the richness of semantics while also not overstretching the user. It was estimated that tagging the videos should have taken about 20-40 minutes.

Existing media

Then the users were asked to search for one or more videos in the system and tag and/or modify these. Finally, users were asked to search for certain keywords in the system to find related videos and fix existing metadata. In order to provide the videos for these two tasks, two videos in each category were uploaded and modelled by the researcher. The keywords would point to an exact video that had been modelled erroneously by the researcher, in order to observe how users would correct metadata.

User groups

The tasks undertaken by each user group is described in this section. Each group uploaded and tagged one or two video streams and one audio stream as previously described in Appendix I. Table III.1 illustrates the relationships between the content categories, user groups and content features.

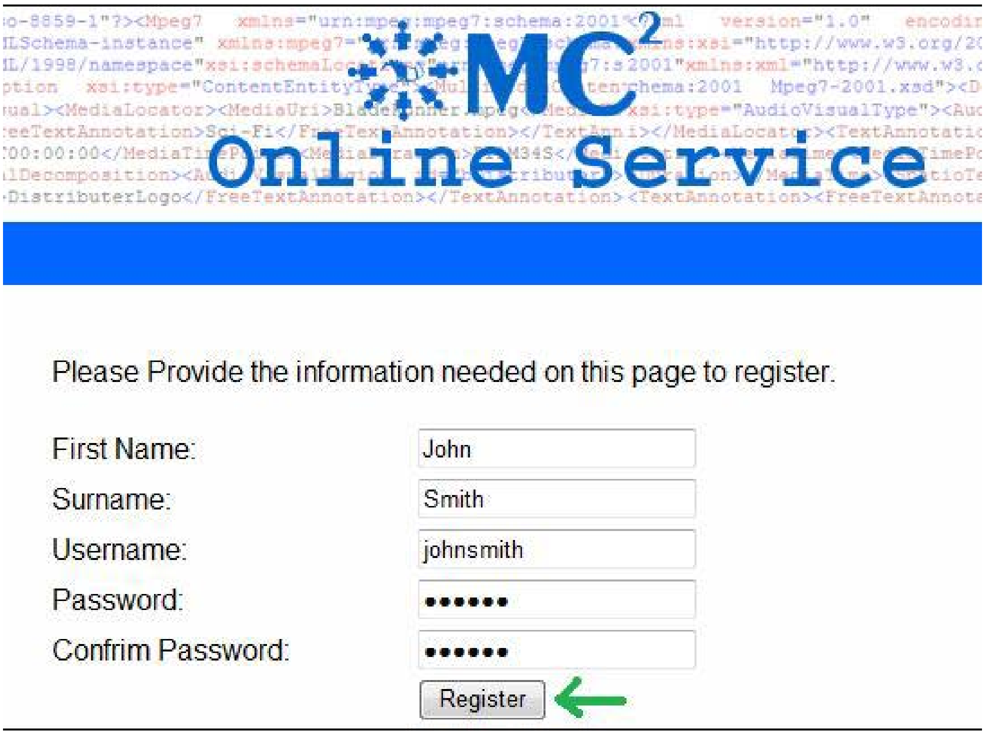
Table III.1. Mapping of content categories to user groups to content features.

	Personal	Business	Academic	Recreation
	User Group 1	**User Group 2**	**User Group 3**	**User Group 4**
People	X	X	X	X
Inanimate Objects	X	X	X	X
Animals	X	X	X	X
Properties	X	X	X	X
Events	X	X	X	X
Music	X			X
Noise	X		X	X
Conversation		X		X
Temporal Relations			X	
Spatial Relations	X		X	
Causer Relations			X	X
User Relations		X	X	
Part Relations		X	X	X
Specializes Relations		X	X	
Location Relations	X			X

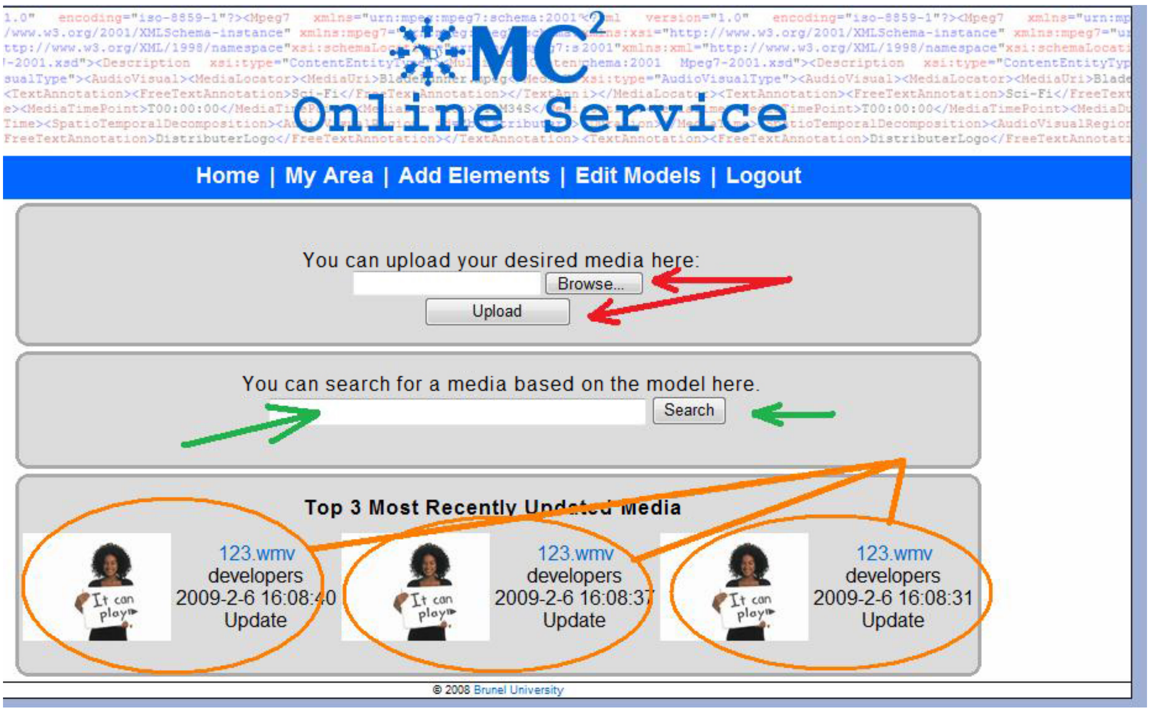
Appendix IV: MC² Online Service User Guide

1. Follow the link from the participants' page so that you arrive on the main service page. Here the service will ask to login or register with the system. If you have already registered just put in your username and password and login. Otherwise click on the link to sign-up.

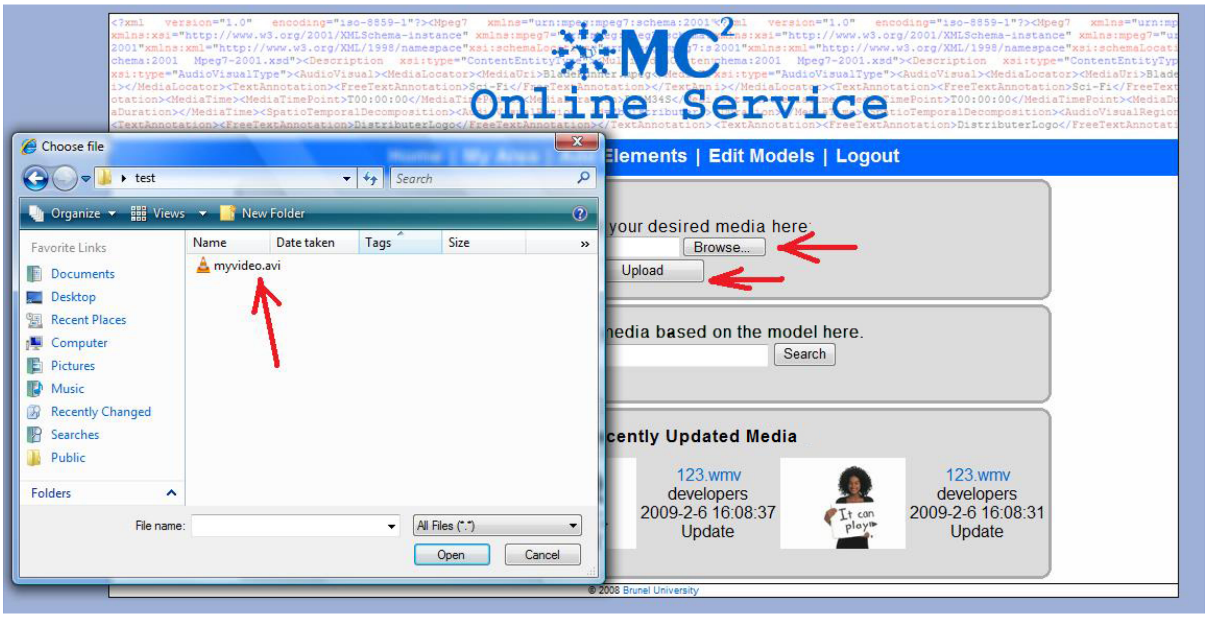
2. The signup page will ask you for your first name and surname, a username of your choice and a password. Please use your real name so that the researchers can identify you and put your data in the right section.



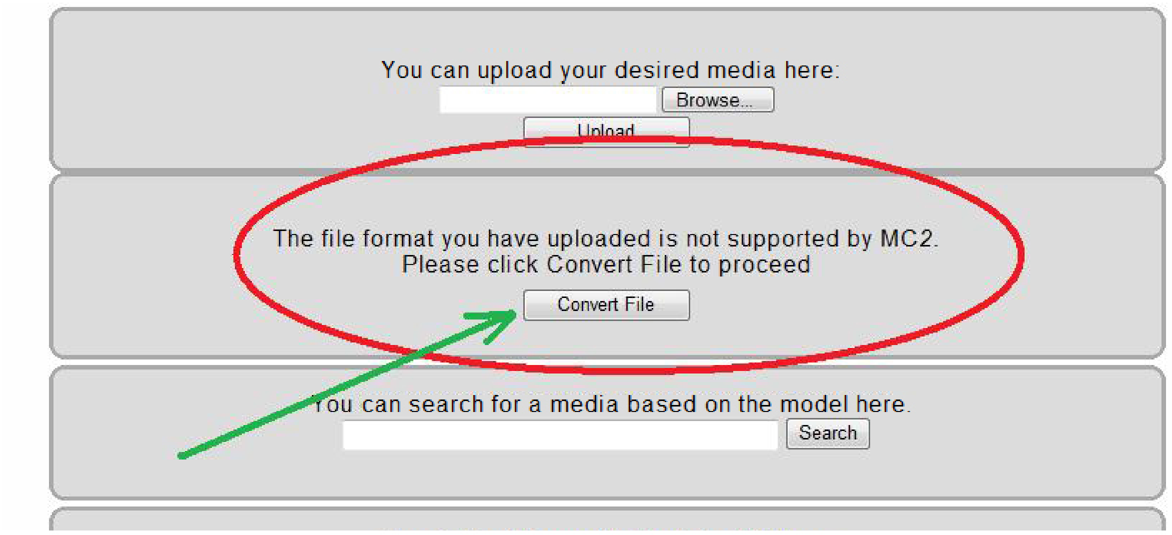
3. After you log in you will be taken to the home page where you can upload a new file, search for an existing file or choose one of the top new files on the system. You can upload a new file by browsing for it on your machine and pressing Upload (red arrows). You can also search for a keyword in the system and press search. Alternatively you can choose one of the top 3 most recently modelled AV files.



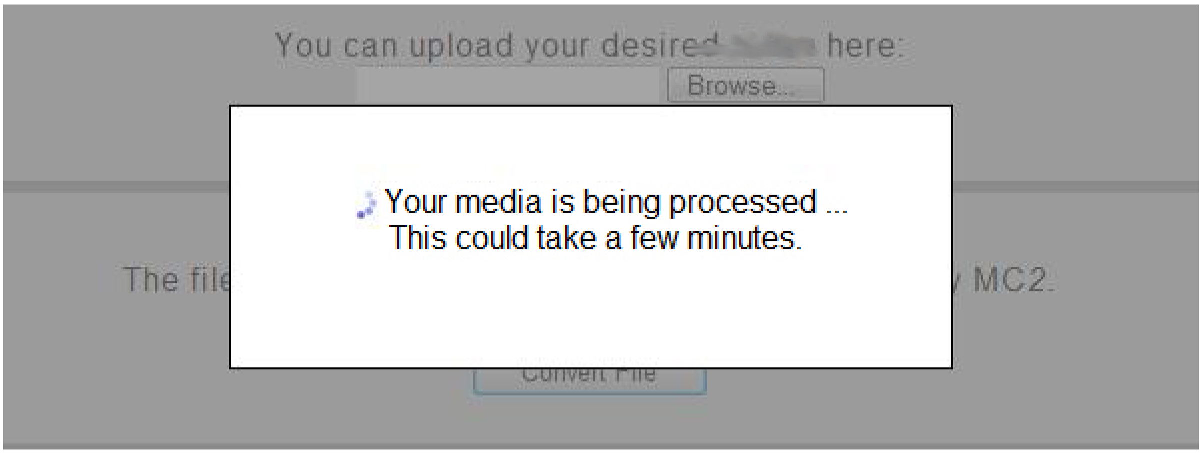
4. If you choose to upload, browse for your file in your system, press open and click on Upload.



5. When your file is uploaded, if it is not in the format that the system accepts, you will be asked to convert it. Just click on the convert button and wait.

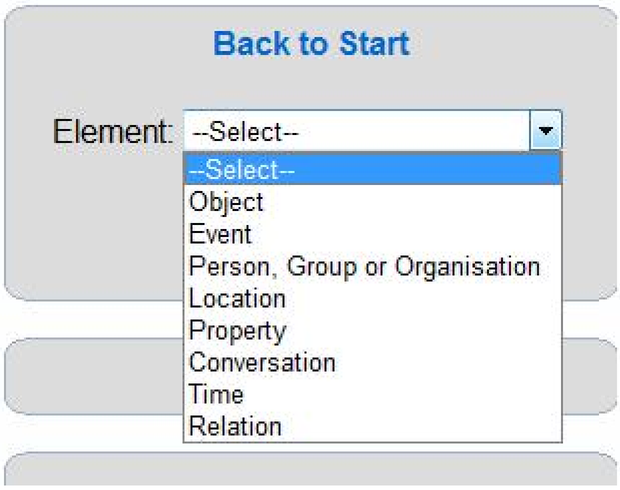


6. While your AV file is converted, you will see the following screen, please be patient and allow the service to finish its task. Please refrain from closing the browser in the midst of the conversion process as your experiment data may become unusable.



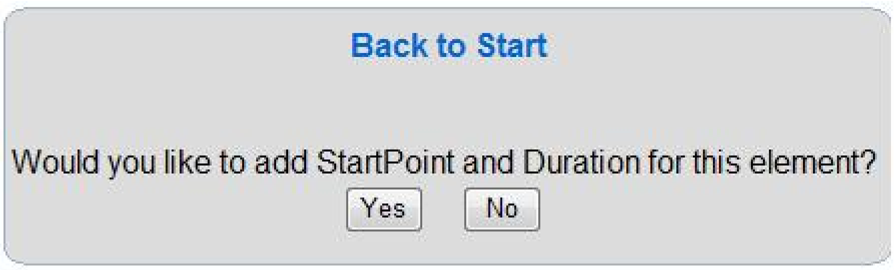
7. When your AV file is converted you will automatically be taken to the screen to begin viewing and modelling it. On this screen you can playback and view your media, model it and add elements to it. You can also view the elements already added to the media. You may also comment on the media, and finally you can use quick add option to quickly model and insert tags.

8. The element dropdown includes all the elements available for adding to the system.



9. You can select an object; in that case you will need to choose a name for it.

10. When you press add you will be asked if you would like to add timepoints for this object

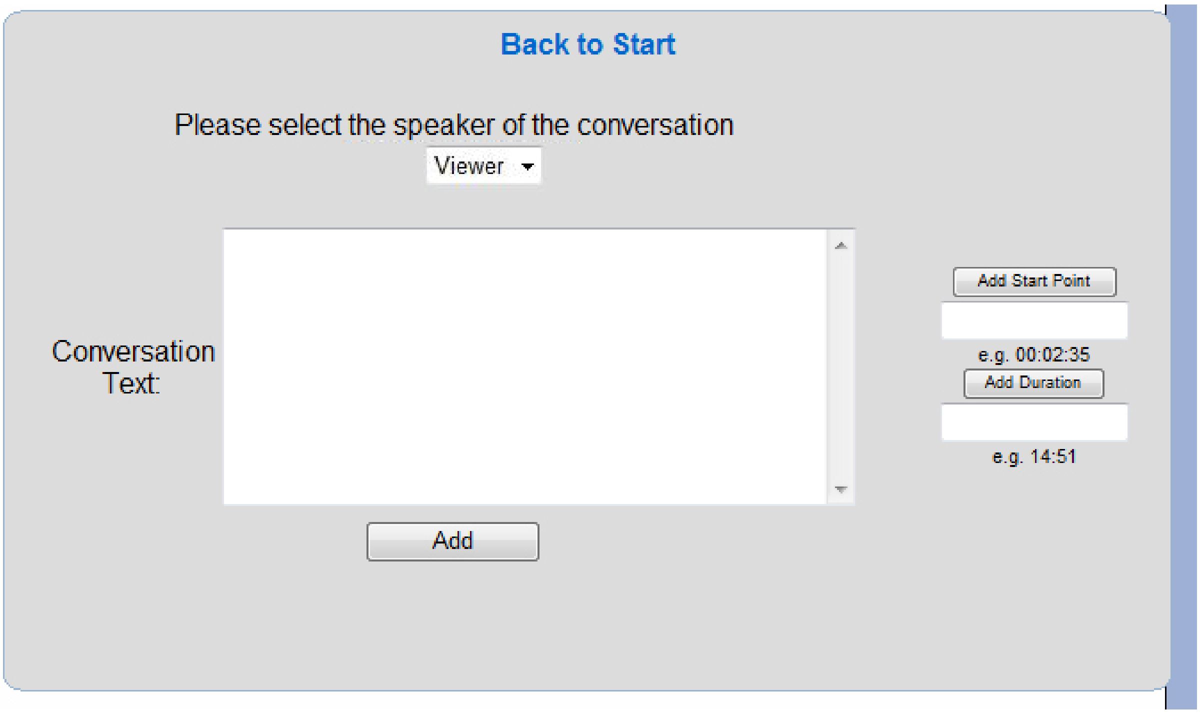


11. If you clicked Yes, you will see new options appearing. **Note that the object is not added to the system yet.** You can enter the point and duration manually or you can browse to the appropriate times on the player and then click on “add start point“ or “add duration” buttons.

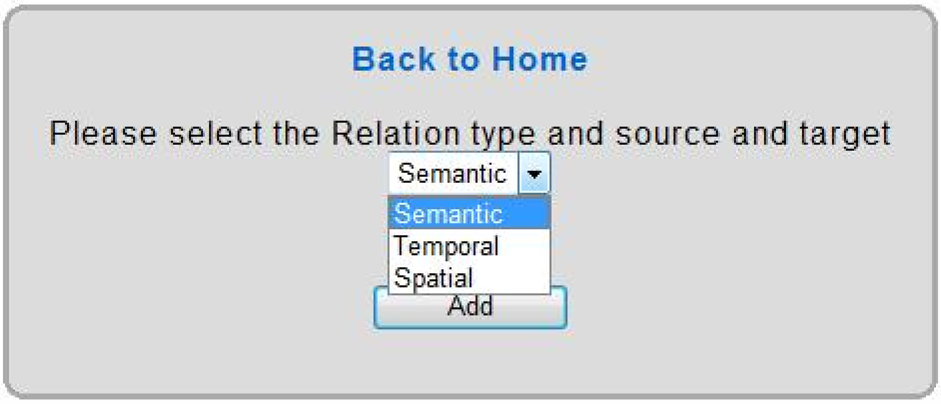
12. From the main menu you can also add properties. A property has a name, value and owner.

    For more information please read the definitions.

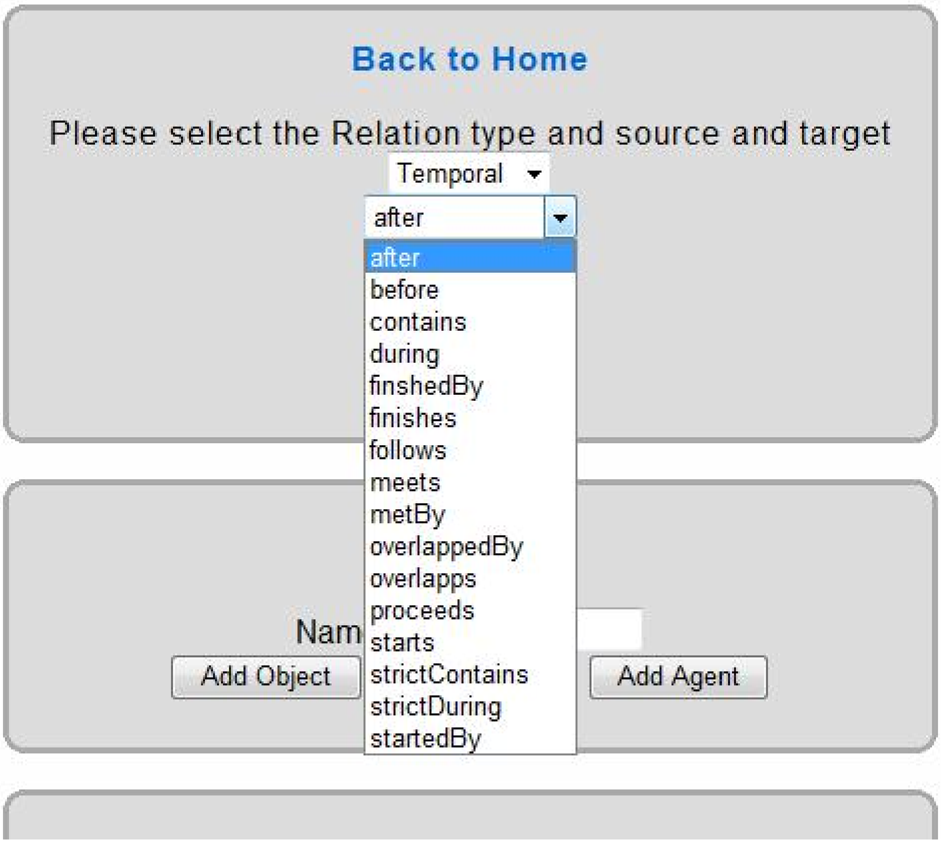
13. You may also add a conversation. The conversation will need a speaker, text and timepoints.



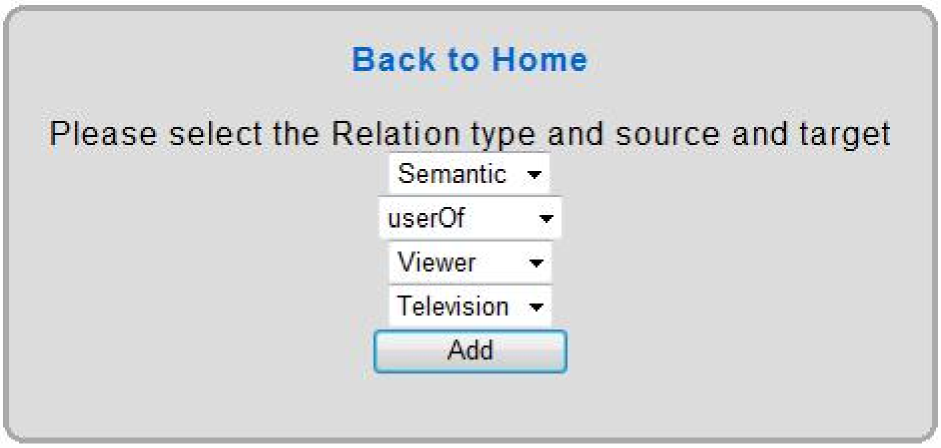
14. You can also add relations. There are three types of relations: semantic, temporal and spatial. Please refer to the definitions for more information.



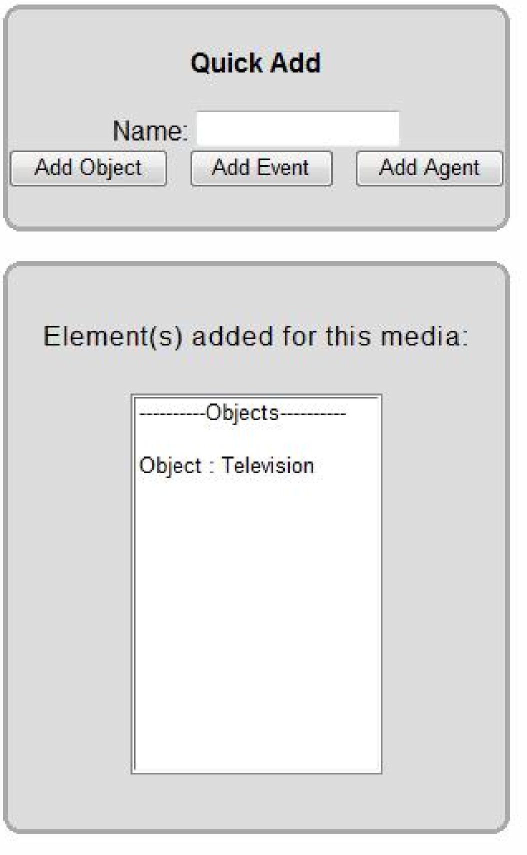
15. Each type of relation will provide you with many options.



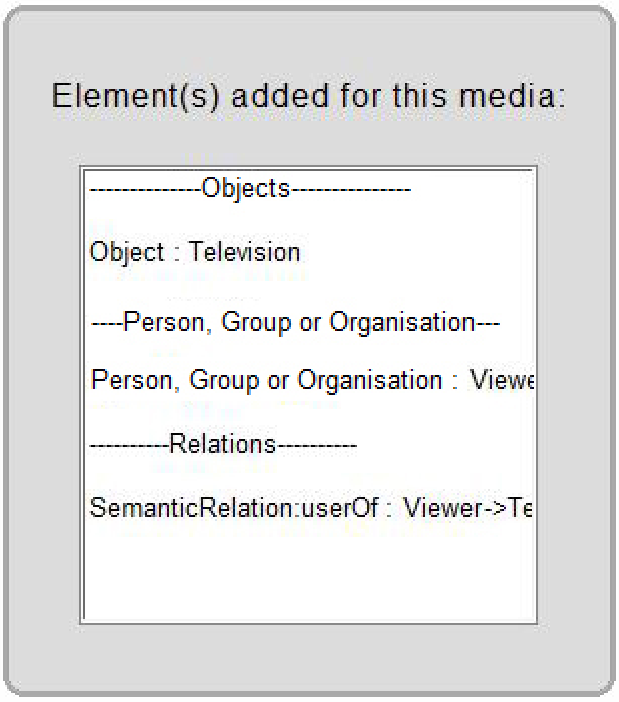
16. To add a relation you will need a source and a target for it.



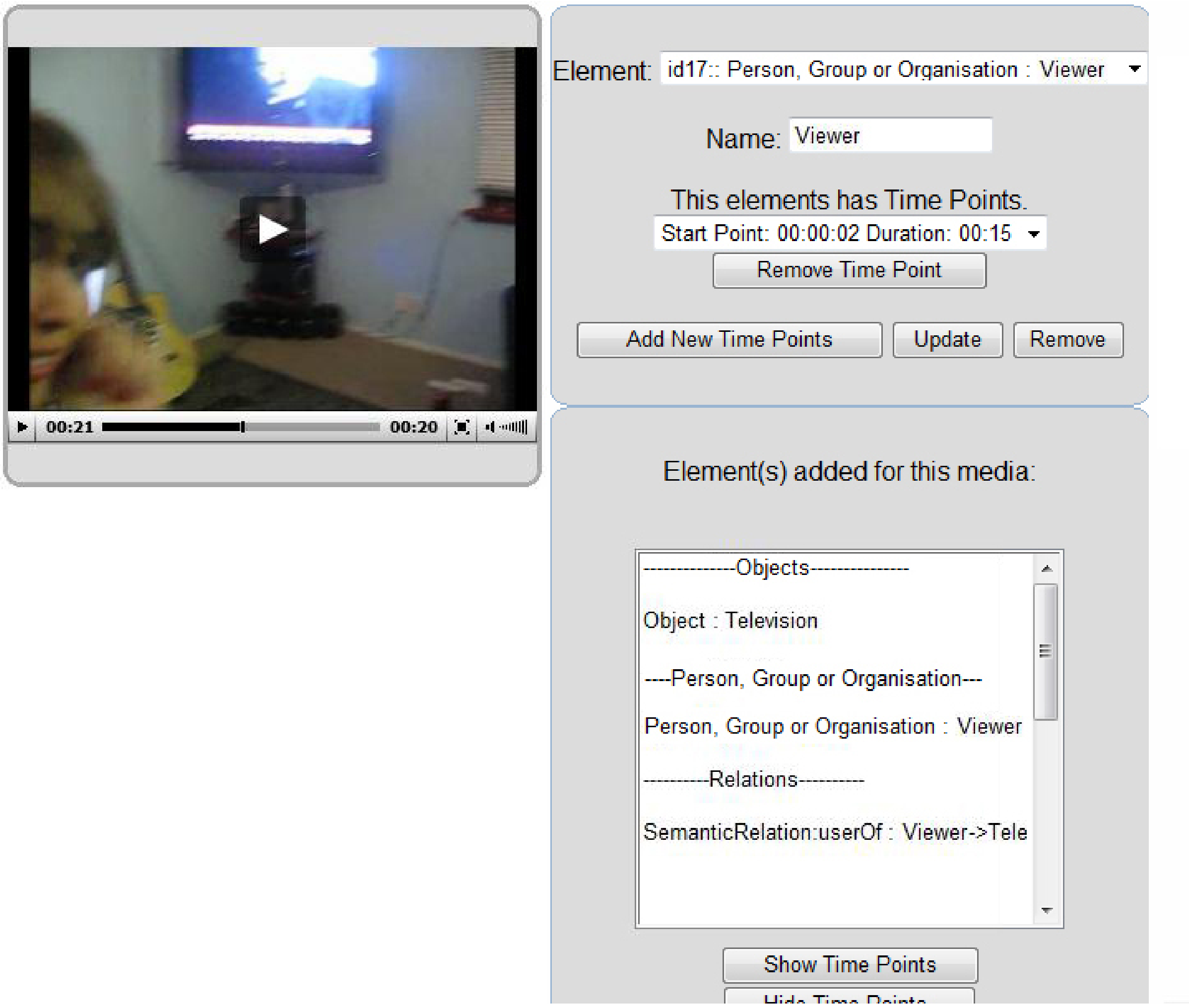
17. You also have the option of using quick add by clicking on the link below the main adding menu which will open the quick add menu for you.



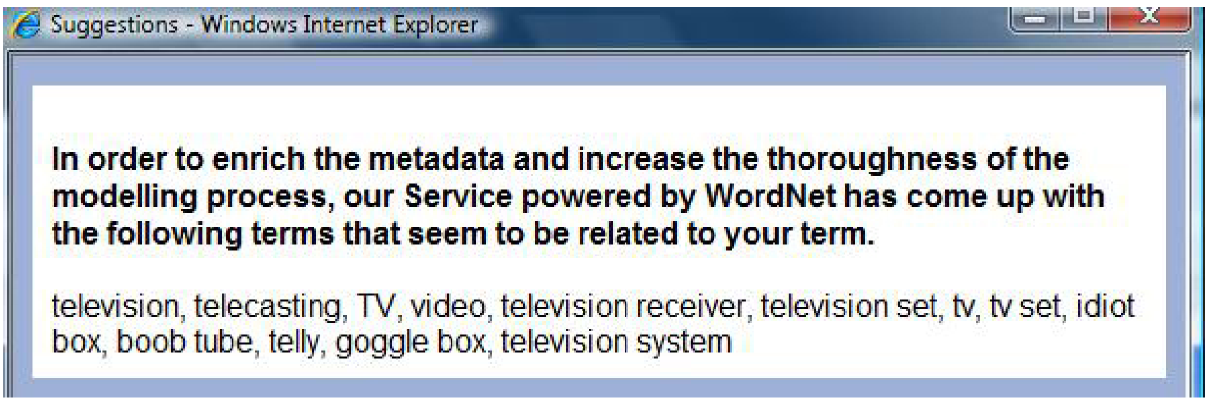
18. The elements you add for the media are shown to you in the elements list box.



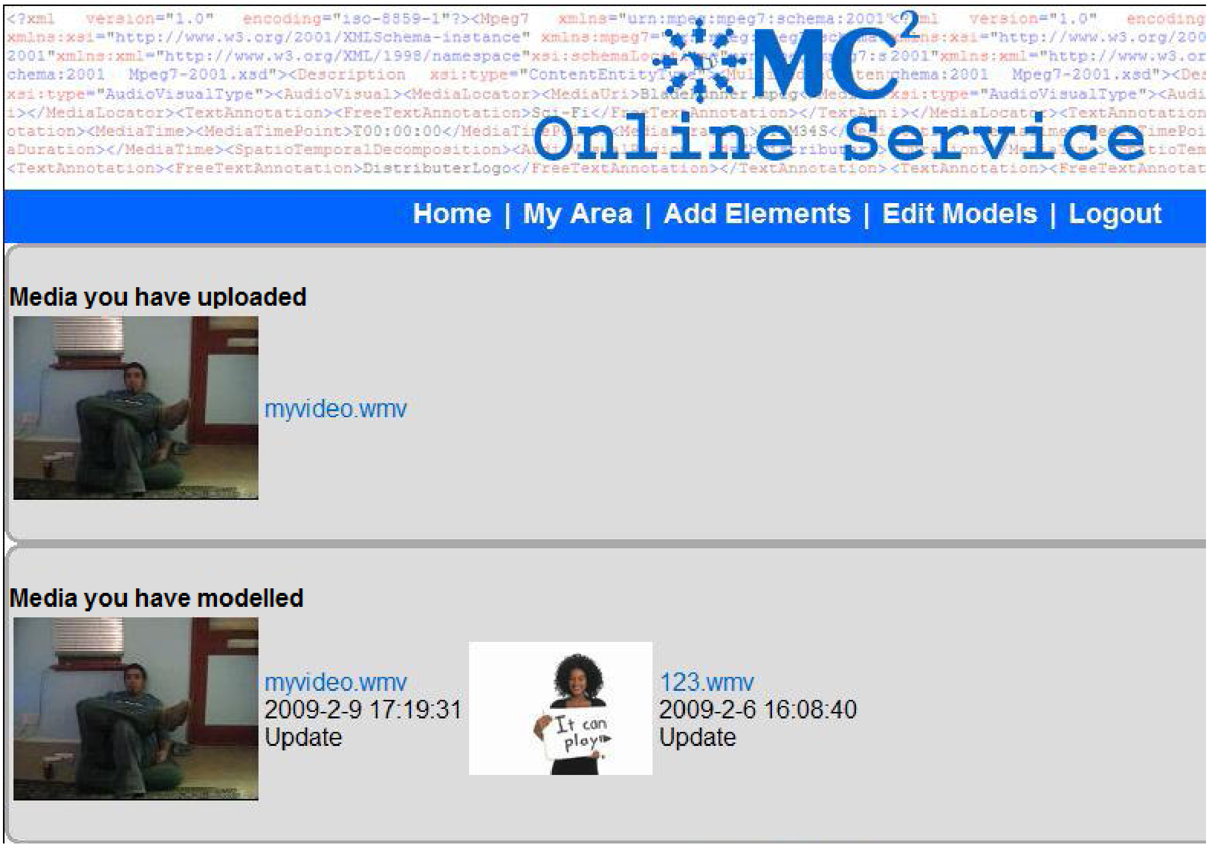
19. You may edit your model or other users' models by going to the edit model page, where you can update or remove elements by selecting them.



20. When a new element is added, the service will suggest similar terms that you can also tag to enrich the metadata.



21. You can view the media you have uploaded or modelled in the My Media page.



22. To logout of the system, simply click on the logout option in the menu bar.

Appendix V: Interview Structure

Initial Experiment Interview Structure

Each interview took approximately 20-30 minutes to complete. It was a semi-structured interview session and the questions and discussion were focused on the following:

1. Of all the tools used, which did you find the easiest to use? Why, why not?
2. Of all the tools used, which did you find the hardest to use? Why, why not?
3. Of all the tools used, which did you find the most functional to use? Why, why not?
4. Of all the tools used, which did you find the least functional to use? Why, why not?
5. Which tags/types of tags did you use most when tagging and describing the AV content? Why?
6. Which tags/types of tags did you use least? Why?
7. Which aspects did you find easiest/use most when tagging and describing the AV content? Why?
8. Which aspects did you find hardest/use least when tagging and describing the AV content? Why?
9. Which aspects do you feel are most important when tagging AV content? Are certain aspects more important for different types of content? Why? Which are least important? (Maybe ask them to rank the aspects.)
10. Did you find any AV files particularly difficult/easy to tag? Why?
11. Did you feel working in a group helped your tagging/description? What about the conversation, did that help? Why or why not?
12. What features would you like to have seen in the tools to help you with your tagging/description?

Repeat experiment Interview Structure

Each interview took approximately 20-60 minutes to complete. It was a semi-structured interview session. The questions and discussion focused on the following:

Tools

1. Rank the three tools used in Experiment 1 and MC^2 from easiest to hardest to use. Discuss why particular tools were easy or hard to use.
2. Rank the three tools from most to least functional. Discuss why particular tools were functional or lacking in functionality. What functionality was missing?

Features/Tags

1. Was it easy to relate what you wanted to describe to the named features? Explain why.
2. Rank the following features according to how **important** they are to:
    a. Videos in your category (explain why)
    b. Audios in your category (explain why)

- People
- Inanimate Objects
- Animals
- Object properties
- Events
- Music
- Noise

- Conversation
- Temporal Relations
- Spatial Relations
- Causer Relations
- User Relations
- Part Relations
- Specializes Relations
- Locations

3. Rank the above features according to how **easy** they were to use for a and b (explain why).
4. Rank the above features according to how **often** they were used for a and b (explain why).
5. Were any additional features required?
6. Did you find any AV files particularly difficult/easy to model? Which ones? Why?

Collaboration/community

1. Did working in a group help with or hinder the process of modelling the AV content? Explain why.
2. Did the conversation via the forum and private messaging help? Why or why not?
3. How satisfied were you overall with the collaboration features in MC^2? Explain why.
4. Would you add or remove any collaboration features? Explain why.

Appendix VI: Sample Interview Transcripts

Sample Interview Transcript 1

1. easiest: youtube, reasons: familiar

2. hardest: flickr, reasons: had to go to pages you didn't need to go, unnecessary interactions

3. most functional: cosmosis, reason: more options, logical structure

4. least functional: flickr, reason: too vague, unclear categories, what type of tags? were all the same

5. most used tags: objects, properties, reason: you think about what you see at the time and what strikes you first

6. least used tags: relations between objects, temporal and spatial, reason: not important

7. easy aspects: objects, properties, events, topic (genre)

8. hard aspects: spatial and temporal relation, reason: not useful

9. most important: high level, topic, objects, properties, people, location. least important: spatial, temporal relations

10. effectiveness: reason: content not interesting, visual stimulates

easy: sports, factual content

hard: movies: reason: narrative based, too many details

11. group: yes, inform your choices better, other points of views, mutual agreements. conversation: no, forum is better, waste of time, missing stuff, distracting

12. existing tags for similar content. viewing tags of similarly minded people. details and overviews of groups and people

Sample Interview Transcript 2

1. easiest to use: flickr, reasons: more familiar, used to it

2. hardest to use: COSMOSIS, reasons: bad interface, too complicated, too many options

3. most functional: cosmosis, reason: too many functions

4. least functional: flickr, reason: limited options, freedom of tagging

5. most used tags: location, object, event, date, people

6. least used: relations, reason: I want to keep it simple

7. aspects: object, people and events, reason: these are what make a video or image

8. relations, spatial and temporal, reason: too many, insignificant, hard to describe, properties, unimportant

9. locations, objects and dates, events

10. eventful ones are easy, reason: too many stuff to tag

11. group: yes, you might miss stuff and they can help. Conversation: no, very distracting can easily go out of hand

12. different types of tags (sports, news, etc..) the categories match the search categories for later, have definitions accessible

Appendix VII: Sample Communication Transcript

beautiful synthesis says (22:56): ok how are we doing this

beautiful synthesis says (22:57): and my video is about cooking not aeroplanes...what are you watching kash?

beautiful synthesis says (22:58): ok here

Kash says (22:58): I am watching people jumping out of planes

beautiful synthesis says (22:58): sounds fun

pwned says (22:59): why hasn't; it finished? >.<

beautiful synthesis says (22:59): its a vaio!!! it must be your fault

beautiful synthesis says (23:00): http://mc2.brunel.ac.uk/Files/video41.mpeg

Kash says (23:00): lol omg k

Kash says (23:00): ok guys You cannot start File Transfer with more than one contact in a conversation.

Kash says (23:00): we are gona upload it on youtube

pwned says (23:03): skydiving fun

Kash says (23:03): hmm

Kash says (23:03): ok so we have to have a plan

beautiful synthesis says (23:03)::/

Kash says (23:03): I am gonna make bullet points

pwned says (23:03): why? improvise

pwned says (23:03): its the way forward, follow your instincts

Kash says (23:04): if I had done that I would have never made it this far

pwned says (23:04): imagine adding base to that video kash, BOOM

beautiful synthesis says (23:04): bullet points for those whose english isn't quite so good

Kash says (23:04): lk

Kash says (23:10): Chi

Kash says (23:10): have you listened to the end of the video

pwned says (23:10): its hard to make out much conversation

Kash says (23:11): hmm

Kash says (23:16): my upload is now complete

Kash says (23:17): what are your youtube usernames? </3

Kash says (23:18): ??

beautiful synthesis says (23:18): motif80 is mine

Kash says (23:23): I am moving to task 2

Kash says (23:23): because I have finished task 1

pwned says (23:23): did u write down u saw sheep?

pwned says (23:23): cos i see white blobs

Kash says (23:23): due to being great and having a good laptop

beautiful synthesis says (23:23): no doubt have finished..

beautiful synthesis says (23:23): oopps *half

beautiful synthesis says (23:24): do you want me to correct your spelling?

Kash says (23:42): ? Kash just sent you a nudge.

pwned says (23:42): second batch

pwned says (23:42): first one Kash just sent you a nudge. You have just sent a nudge.

pwned says (23:43): did u tag the second batch

pwned says (23:43): or just descriptions

Kash says (23:43): i tag only so far

Kash says (23:43): the manual is not very good is it

pwned says (23:43): nope

pwned says (23:43): ask

Kash says (23:43): tells you something about the person who wrote it

pwned says (23:43): lol

pwned says (23:45): did u get all football players?

Kash says (23:45): no

Kash says (23:45): i got bikes

Kash says (23:46): omg k

Kash says (23:47): oooooooooh

pwned says (00:14): no

pwned says (00:14): not

pwned says (00:14): yet

pwned says (00:55): Download COSMOSIS from here and consult the section regarding COSMOSIS in the user guides PDF.

pwned says (00:55): wheres the pdf?

pwned says (01:01): ur hardlink

pwned says (01:46): well thats a major problem in this world

pwned says (01:46): yes

pwned says (01:46): it is

Glossary of Terms

4oD: The Video On Demand service of Channel Four Television Corporation, UK, that offers a the programmes previously shown on Channel 4, E4 and More4.

AJAX: (Asynchronous JavaScript and XML) The group of techniques for web development that allow web pages to be updated asynchronously.

ASP.NET: A Micorsoft server-side dynamic Web page application development framework.

AV: Audio-Visual.

BBC iPlayer: BBC's Video and Audio on Demand service for programmes previously shown on BBC TV channels.

COSMOSIS: An MPEG-7 system that supports extensive individual tagging.

CSS: (Cascading Style Sheets) A markup language used for describing the presentation semantics (look and feel) of documents. Its' most common application is to style web pages.

Data clustering: The unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters) based on similarity. It is a main task in exploratory data mining, and a common technique for statistical data analysis.

Del.icio.us: A social bookmarking web service for storing, sharing, and discovering web bookmarks.

Facebook: An online social networking service.

FLAC: A lossless digital audio codec. it allows the file size to be reduced without any information being lost.

Flickr: An image and video hosting website and an online community which is widely used by photo researchers and bloggers to host images that they embed in blogs and social media.

Folksonomy: A classification system derived from the practice and method of collaboratively creating and managing tags for content annotation and categorisation.

HTML: (HyperText Markup Language) The main markup language for creating web pages and other information that can be displayed in a web browser.

Hulu: A website offering on-demand streaming video of TV shows, movies and other new media from major American networks and studios. Hulu videos are currently offered only to users in the United States and its' overseas territories and is blocked for users outside the US using IP address location.

IMDb: (Internet Movie Database) An online database of information related to films, television programs, and video games. This also includes information on actors, production crew personnel, and fictional characters featured in all three visual entertainment media.

LINQ: (Language Integrated Query) A Microsoft .NET Framework component that adds native data querying capability to .NET languages.

Metadata: "data about data". Examples include: tags, creation information and card catalogues of libraries. Metadata are also used to describe digital data using specific standards.

MPEG-4: A method of defining compression of AV digital data.

Netflix: An American provider of on demand Internet streaming media.

Rapid Application Development: A software development methodology that uses minimal planning in favour of rapid prototyping. Software development planning using RAD is interleaved with software programming. The lack of extensive pre-planning generally allows software to be written much faster, and makes it easier to change requirements. RAD is not appropriate when technical risks are high.

Sky Anytime+: BSkyB's video on demand service for Sky customers within the UK.

UML: (Unified Modeling Language) A standardised (ISO/IEC 19501:2005) and general-purpose modeling language for software engineering. It includes a set of graphic notation techniques to create visual models of object-oriented software-intensive systems.

Use Case: A use case is a list of steps, typically defining interactions between a role (known as an "actor" in UML) and a system to achieve a goal.

Vorbis: A free and open-source software project which produces an audio format specification and software implementation (codec) for lossy audio compression.

WavPack: A free, open source lossless audio compression format.

Web 2.0: Describes web sites that use technology beyond the static pages of earlier web sites. A Web 2.0 site may allow users to interact and collaborate with each other in a social media environment as creators of user-generated content.

Windows Media Player: A media player and library application developed by Microsoft that is used for playing audio, video and viewing images.

WMP Tag Plus: A plug-in that integrates seamlessly into Windows Media Player to provide library and tagging support for additional music formats. These formats include FLAC, Ogg Vorbis, WavPack, Monkey's Audio, Musepack and MPEG-4.

WordNet: An English semantic lexicon database, created at the Cognitive Science Laboratory of Princeton University.

XML: (Extensible Markup Language) A markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

YouTube: A video-sharing website on which users can upload, view and share videos.