

Matching Demand and Offer in On-line Provision: a Longitudinal Study of Monster.com

Andrea Capiluppi and Andres Baravalle
School of Computing, Information Technology and Engineering
University of East London, UK
{a.capiluppi,a.baravalle}@uel.ac.uk

Abstract—When considering the jobs market, changes or recurring trends for skilled employees expressed by employers’ needs have a tremendous impact on the evolution of website content. On-line jobs sites adverts, academic institutions and professional development “standard bodies” all share those needs as their common driver for contents evolution.

This paper aims, on one hand, to discuss and to analyse how current needs and requirements (“demand”) of IT skills in the UK job market drive the contents of different types of websites, in turn analysing whether this demand changes and how. On the other hand, it is studied what the UK higher education institutions have to offer to fulfill this demand.

The results found analysing the evolution of the largest on-line job centre (www.monster.com), and the websites of selected UK academic institutions, demonstrate that often what is requested by UK industries is not clearly offered by UK institutions. Given the prominence of monster.com in the global economy, these results could provide a meaningful starting point to support curricula development in UK, as much as worldwide.

I. INTRODUCTION

Part of the role of professional bodies and higher education institutions is, today as from their earlier times, to prepare students for a productive role in the society. Universities, colleges, the wider range of course providers, and standard bodies face the responsibility to effectively interface with the jobs market, by trying to understand what are the skills that are missing in the market and to react on how those can be delivered. In the UK, this picture is becoming ever more complicated, since the economy is in deep crisis and unemployment is soaring: unemployment in the age band 18-24 has reached 746,000 in the period Jul-Sep 2009 [1].

On the one hand, the SFIAPlus framework [2], developed by the British Computing Society (BCS) over the past 20 years in UK, has already been at least partially successful in identifying the skills that are needed by UK companies. As a starting point of the SFIAPlus framework there are the requirements of UK IT companies, as expressed by them and collected by the BCS with peer discussion groups, interviews, panels etc over the years. The result is a list of high level skills, hierarchically organised that can be used by employers, professionals, educators for their different aims. Employers can use the SFIAPlus framework to create job profiles (e.g. when recruiting); educators can use it to describe their academic programmes, and professionals can use it to match their skills to market requests, and to plan their development path. While the importance of the SFIAPlus framework is

not under discussion, this paper aims to analyse the problem with a finer granularity. The first objective of this work is to show that the jobs market is not typically looking for someone with generic “Software development” skills (one of the SFIAPlus categories), but for *specific* skills such as Java, .NET or C++ developers (just to give an example). This creates a short circuit between “what is needed” and “how it should be described”, that at least partially undermines the usefulness of such standard bodies, apart from their responsiveness to changing needs.

On the other hand, higher education institutions and “generic” course providers must ensure that students are able to acquire skills that are in demand by the jobs market, both by providing relevant and needed skills, and by further developing their curricula. The latter is of uttermost importance in areas as IT, where wrong choices by students and/or weak provisions by course providers might have repercussions in the employability at the end of the studies. According to the Universities & Colleges Admissions Service (UCAS), a total of 477,277 students had a place confirmed in a UK university in 2009 [3], 18,239 (3.8%) of which were in the categories Computer Science (11,328), Information Systems (3,165), combinations of Mathematics and Computer Science (2,229), Software Engineering (1,466), and Artificial Intelligence (51). 81 UK universities are listed in 2009-10 as providers of either undergraduate or postgraduate (or both) IT courses. What needs to be clarified is whether this “offer” is aligned with the jobs “demand”: the second objective of this paper is therefore to study whether academic institutions in UK can effectively show what their courses offer, and whether this offer is in-line with what is requested by the market. To counteract these effects, it has been reported that institutions of higher education and IT standardisation bodies are used to redesigning their curriculum, but only once every four or five years [4].

In the proposed scenario, this paper is an empirical attempt to analyse the two sides of demand and offer of IT skills in the UK market: this problem has emerged before in the literature, and to the best of our knowledge it has just been pointed at, but not solved ([5], [6]). On one side, this paper will discuss and analyse what are the requirements of the job market, now, today, for IT workers in UK. An in-depth analysis of the IT job market will be run during the past 9 months (from September 2009 to May 2010), by regularly

parsing the dynamic contents of the largest repository of on-line jobs available in UK (<http://www.monster.com>). On the other side, the on-line course descriptions of selected academic institutions will be analysed, in order to discover what they offer, and to decide whether their offer matches the jobs demand.

This paper is articulated as follows: section II introduces the goal and research questions of this study, and describes how the contents of monster.com were parsed, and how the on-line provisions of selected universities were analysed. Section III illustrates the results of the longitudinal study on the contents of monster.com, while section IV summarises the findings of parsing the contents of all the courses offered by a selection of 10 UK academic institutions. Since the study is based on gathering empirical data, section VI describes some of the associated threats to validity, and how remedial actions were taken, while section VII concludes and proposes some further avenues of development.

II. METHOD

This section introduces the definitions used in the following empirical study and presents the general objective of this work, and it does that in the formal way proposed by the *Goal-Question-Metric* (GQM) framework [7]. The GQM approach evaluates whether a goal has been reached, by associating that goal with questions that explain it from an operational point of view, and providing the basis for applying metrics to answer these questions. This study follows this approach by developing, from the wider goal of this research, the necessary questions to address the goal and then determining the metrics necessary for answering the questions.

Goal: the main aims of this research are first to provide a tool, and a methodology, to continuously report on the current needs formulated by (any) jobs market; and second, to regularly monitor standardisation bodies, academic institutions and course providers to assess the matching of their provisions to those needs.

Question: in this paper, the following research questions have been evaluated:

- 1) Is it possible to quantify and compare the responsiveness to change in the evolution of the websites related to the “demand” and “offer” of IT skill-sets?
- 2) Are there regularities among the IT keywords and required skills (as in recurring job requirements) when considering the evolution of contents within the monster.com website?
- 3) Are the websites of UK academic institutions embracing a common framework to display the on-line contents of their courses?
- 4) Do UK academic institutions attempt to use relevant and recurring keywords, as the ones found on monster.com, when describing their on-line provision of courses?

Metrics: the evolutionary analysis of the monster.com website consists of the extraction of the most relevant “keywords” appearing in job descriptions related only to the IT skill-set. These keywords will be both monitored in time, and searched

in the on-line descriptions of academic courses offered by UK universities. Furthermore, the density of such keywords in the web-pages containing the course descriptions will also be extracted.

A. Analysis of demand – Empirical Approach

The first part of this research had to deal with identifying the data sources representing the “demand” of IT jobs in this study. We have decided from the beginning that we were going to consider only job listings that were featured on on-line dedicated search engines. Focusing only on these types of adverts allowed us to be able to create an extensive data bank and to have significant quantitative data to analyse.

While there are many employment web sites, no definitive study has been conducted on their respective market share. With regards to the US market, according to comScore, a specialist in measuring the digital word, the web sites which catch most of the traffic (in order of importance) are CareerBuilder.com, Monster.com and Yahoo! HotJobs [8]. There are no comparable existing studies of the U.K. market share of the different employment web sites. A preliminary study focused on the identification of appropriate sources for UK, with this ranking:

- 1) CareerBuilder.com,
- 2) Monster.com
- 3) Yahoo! HotJobs
- 4) the Guardian website
- 5) the Times website

CareerBuilder.com was immediately discarded as it is mostly focused on the US market. Yahoo! HotJobs, while addressing the UK market, was discarded for technical reasons, since it does not have a separate IT category, but only a far more inclusive “Technology” category with no subcategories. Moreover, Yahoo! HotJobs does not have an internal archive of jobs, but indexes external job listings; the implication is that the jobs are in very different formats and it would have been challenging to create generic technology that would have been able to analyse such different sites. The other major players in the UK on-line job advertisement market include the Guardian (about 150 IT jobs) and the Times (about 70 jobs).

Monster.com is at the first or second position in the world for number of users and number of jobs posted (depending on the sources used). Powered by Open Source library Lucene, at the time of writing, Monster had 150 million resumes and over 63 million job seekers per month (Lucid Imagination, n.d.). About 4,000 IT jobs were simultaneously on-line in Monster UK at the time of our research, which made it the most important player in the market under evaluation.

B. Analysis of demand – Tool Infrastructure

Once the source for our data was identified, the next step was the data collection. To proceed, we decided to develop customised analysis software, to allow us to automatise the data collection: since it has been reported that building a web crawler is relatively simple, and a comparison of crawlers is outside the scope of this paper [9], an in-house prototype

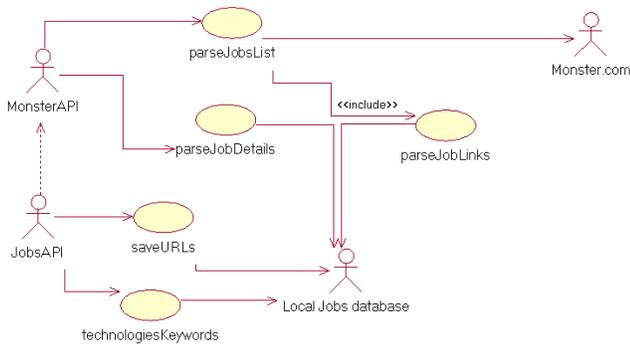


Fig. 1. Use case diagram – monster.com spider

was developed. At this stage, it is worth noting that we decided from the beginning not to look for jobs in the web in general, as we felt that the advantages of such approach were outweighed by the exponentially higher costs in terms of software development and by the computation resources for data mining. The software developed for this research includes an ad-hoc **web spider**, a **data extractor** module and a basic **entity recognizer** component (see Figure 1).

While only the monster.com spider has been fully developed, the system has been developed with extensions in mind and further modules can be plugged in if required or desired. The spider is in charge of downloading all the raw data from Monster.com. Jobs in Monster.com are organised in hierarchical categories, and only the jobs in the IT categories have been downloaded. The limited processing power of our equipment and compliance with netiquette rules led us to decide to minimise the interaction with Monster.com and the overload to their resources.

After a trial and error phase, we identified a suitable number of “runs” per day: the software runs 7 times per day looking for new jobs, and 4 times per day re-parsing the full catalogue. Running the spider several times per day allows us to minimise the risk of missing any job listings, and to be able to obtain timely dates for the removal of the jobs.

The **data extractor** is in charge of processing and categorising the raw data. Its first function is to identify and discard corrupted or incomplete data (e.g. “dummy” or “test” job listings), and then to extract the most relevant information: job title, company, location, industry, type of job, career level, job URL, salary and description. This is done using regular expressions.

The final component, the **entity extractor**, extracts a list of Information Technologies from the job listings. A set of regular expressions and a gazette are used to extract IT technologies names from job listings. The spider, the data extractor and the entity recognizer all run through cron jobs, several times per day, to minimise the load on Monster.com and on the server who hosts the scripts itself. A set of scripts, running against the data bank, are used as a final filter to extract the relevant information. A human analysis is then

performed to ensure the correct identification of IT terms and to group the job listings in appropriate categories.

C. Analysis of Offer – Empirical Approach

The second part of this research dealt with identifying the data sources representing the on-line “offer” (or provision) of IT skills in this study. A list of higher education institutions offering IT- and computer science-related under- or post-graduate degrees is available on the Guardian website¹. This same list contains the number of currently registered students: based on the number of students, the top 10 universities were selected for a further analysis of their on-line course contents. This initial selection was obtained because it is arguable that a larger number of students requires an increasing need of accurate on-line descriptors of the provided courses.

The subsequent data gathering phase considered all the courses (e.g. BSc, BSc Hons, etc) and all individual modules (e.g. “Introduction to Programming”, “Databases I”, etc) offered by this selection of academic institutions, but only at the undergraduate level: this was needed given the uneven distribution in the number of BSc and MSc courses at each selected university. In summary, for each academic institution, the process of gathering the contents of courses consisted in:

- 1) first, a manual identification is needed in order to list the IT- or computer science-related courses, at the undergraduate level, from the web-pages of the selected UK universities;
- 2) secondly, it is necessary to recursive download the courses’ on-line pages, and subsequently to extract the list of modules contained within such courses;
- 3) thirdly, having gained the list of modules as above, the web-pages detailing each module are downloaded;
- 4) finally, from the description of both modules and courses, the paragraphs detailing the *contents* of both courses and modules, and/or the available *skill set* at the end of each course or module are identified and extracted.

Given the variety of resources, the only common part to all the selected universities was the recursive downloading of courses. Obtaining individual modules, and the parsing of contents varied in each case, and had to be conducted via a one-to-one implementation and use of shell or Perl scripts.

III. RESULTS – DEMAND

At this stage, the monster.com spider has been running for just over 8 months, collecting and analysing more than 48,000 jobs in the selected category. Our first step was to analyse the job listings, to look for any trends in the selected time-frame: to do that, we took monthly snapshots of the information. Table I lists 30 IT keywords which appeared most reoccurring in the very first extraction (September 2009), and it tracks their evolutionary relevance in the subsequent 9 months. In Figure 2, the graphical evolution of the first 10 keywords is

¹<http://www.guardian.co.uk/education/table/2010/feb/15/computer-sciences-it-postgraduate-masters-table>

shown (the other 20 expressing proportionally a very similar pattern).

A. Results – IT jobs in UK

The first thing that we can note from Table I is that the request for specific skills has been quite constant in the selected time-frame. Most of the skills did not change position in the listing at all, and the ones that did change position only climbed or fell a few steps.

With regards to the software platforms, there are some clear indicators on the UK market:

- as pointed out in other studies [10], the demand of IT jobs in UK is constantly growing, apart from a monthly decrease (December 2009) where the visible step-wise decrease is arguably connected with the Christmas holidays period;
- Microsoft-specific technologies are predominant in the IT market. The ASP.NET framework is the most requested technology, and .NET and VB.NET are both individually present in the first 30 positions;
- Sun technologies (Java, J2EE) are also featured prominently;
- Linux and other open source skills (MySQL, PHP) are nevertheless one of the most requested skills, and its request (similarly to the others) is going up, proportionally (see Figure 2).

While Microsoft is clearly predominant, the request for such a (comparably) high number of professionals with experience in Linux should be used as an indicator to check whether those skills are in shortage in UK.

From just looking at the numbers, it is quite visible that server side web technologies appear in the top positions: ASP.NET, PHP, J2EE; similarly significant is the fact the order in which they appear. Design-related technologies also feature prominently (HTML, XHTML, Photoshop, ActionScript, Dreamweaver), with HTML as one of the most desired skill. Programming skills are featured across the full list and C++ is specifically featured.

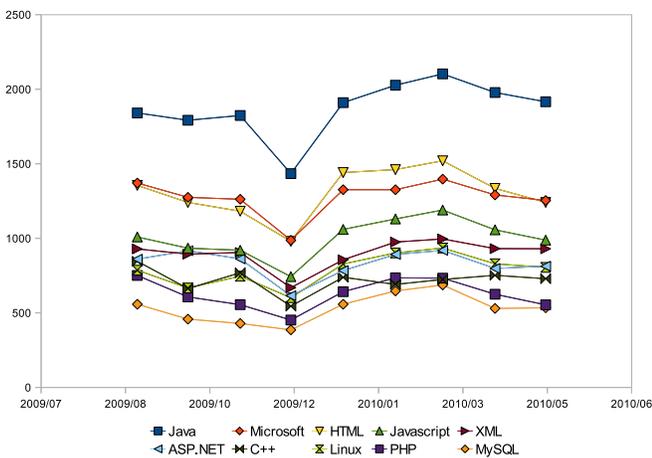


Fig. 2. Evolutionary trends of top 10 Information Technologies skills

B. Results – Server-side web development

The next step was to analyse the performance of different technologies for web development. We have focused our analysis on .NET, PHP and J2EE, again using two snapshots.

Table III shows how Microsoft technologies for server-side web development are clearly more in demand than the other two more common alternatives: PHP and J2EE. At the same time, it shows a number of interesting facts:

- The percentage increase in jobs on Microsoft technologies is the lowest, comparing both with PHP and with J2EE;
- Less preferred combinations of skills are increasing slower than the preferred combinations (e.g. ASP.NET and MySQL and PHP and SQL Server);
- Demand for PHP+Oracle skills has greatly increased. This is likely to be due to an expansion of PHP into sectors where Oracle is used (rather than the other way around).

It is interesting to compare it with Netcrafts survey of Web servers [11]. Apache, after losing market almost continuously since October 2005, has started to regain market positions. IIS has been increasing its market share till October 2008, and decreasing afterwards.

C. Results – Operating Systems

Many IT jobs require experience in one or more operating systems. This is the case for example in system administration or helpdesk positions. This section analyses operating system skills named in the job adverts.

Our research shows that recruiters tend not to specify specific versions of operating systems. Table 4 shows how “Microsoft Windows” and “Linux” are both named in thousands of adverts, while specific versions only in tens or hundreds. With regard to server-oriented operating systems, the following has been observed:

- Generic “Linux” and “unix” skills together are as popular as those on Microsoft Windows. SUSE and Debian are the most popular Linux distributions.
- Solaris is also very popular, and it outperforms all (individual) Linux distributions and Windows Server versions.
- apart from Windows Vista, the various flavours of Windows have a similar share of number of ads.

IV. RESULTS – OFFER

This section summarises the findings related to the analysis of the selected universities websites, by parsing and extracting the relevant keywords from the contents of courses and individual modules. The results as shown below are not intended to “name-and-shame” any academic institution: the actual names of the universities have been removed, and an anonymous ID is used for each (as in Table IV, ordered by number of IT-related courses).

The first finding when analysing the on-line course and module descriptors of UK academic institutions is the variety of formats used: 5 out of 10 only describe the courses that

Keyword	2009/09	2009/10	2009/11	2009/12	2010/01	2010/02	2010/03	2010/04	2010/05
Java	1,841	1,792	1,824	1,434	1,910	2,027	2,103	1,978	1,916
Microsoft	1,371	1,275	1,262	987	1,326	1,326	1,397	1,291	1,254
HTML	1,355	1,241	1,182	980	1,441	1,461	1,520	1,335	1,241
Javascript	1,009	934	920	743	1,060	1,130	1,190	1,057	988
XML	929	894	905	669	856	975	996	931	930
ASP.NET	860	917	862	615	784	892	919	798	813
C++	846	663	769	545	739	691	724	752	729
Linux	786	667	746	602	828	902	935	829	805
PHP	751	607	555	452	642	735	733	625	554
MySQL	558	458	429	386	558	647	687	530	534
.NET	540	422	469	389	558	511	485	440	444
XHTML	403	371	389	335	418	432	429	302	305
ITIL	401	461	386	313	556	470	490	509	411
CRM	340	297	297	209	402	418	427	354	314
Photoshop	339	315	299	205	356	332	395	313	259
SharePoint	332	309	278	218	287	317	303	304	249
VB.NET	304	330	317	223	307	331	353	311	309
J2EE	287	243	281	254	313	304	331	343	339
UML	284	251	249	217	257	259	277	255	292
SOA	241	216	236	174	217	241	268	266	278
TCP/IP	227	285	277	192	267	246	264	249	272
SEO	227	191	194	149	173	219	200	119	163
MCSE	211	163	170	110	180	157	162	203	175
VMware	190	201	198	144	190	224	228	240	237
ActionScript	183	142	152	81	154	140	210	169	131
Citrix	182	171	225	150	213	182	191	198	204
Prince2	175	188	170	134	195	196	177	219	156
Dreamweaver	129	105	149	102	124	116	143	120	82
PLSQL	128	114	164	96	122	136	159	143	133
MVC	124	110	136	89	159	231	226	217	207

TABLE I
TOP 30 INFORMATION TECHNOLOGIES SKILLS BY JOB LISTINGS

prospective students can enrol to, but no further information is given to describe the modules (coded as “n/a” in Table IV). This is of course detrimental for understanding what courses and individual modules offer to students, let alone defining precisely what is the structure of each course. This of course has an effect on the overall text that is available to new students, which ranges between an overall of 170 lines to over 12,000 lines to describe the global IT- and computing-related offerings of one university.

The second finding is based on the number of keywords contained in these descriptions: when using the list reported in table I, it is found that 6 out of 10 universities describe their courses by using less or at most 5 of the keywords most looked after by IT companies. Only 2 universities use 21 and 22 keywords respectively (IDs 1 and 3 respectively), proposing themselves as the most up-to-date in the interface with the IT industry in UK.

The third finding is related to the keyword density: when analysing all the 30 proposed keywords, and the length of the courses and modules descriptors, it is found that, in most cases, these keywords are very sparse in the text. For instance, the university with ID 4 having only one keyword combined with a long (and keyword-unrelated) descriptive text (figure 3, top). Other institutions score very low in the usage of such keywords in the text: the only university scoring high both in the relative use of relevant keywords, and their density, is

university with ID 1.

The fourth finding is that some keywords are used by several universities in their descriptions, others by very few: C++, HTML and Java are mentioned by 6 out of 10 universities, while more recent technologies (e.g., PRINCE2, J2EE) are mentioned by just one university. Furthermore, 6 keywords are not mentioned by any university (Citrix, ITIL, MCSE, SEO, SharePoint, VMware), although all following the same evolutionary patterns than the other cited in table I.

V. RELATED WORK

The topic of evolution of web-site “contents” is not new: various other researchers have addressed the problem of crawling, on a regular basis, the contents of websites to study their evolution, or how existing pages change in time. A massive wealth of some 720,000 pages [12], and some 151 million [13] Web pages have been downloaded and studied for evaluating recurring patterns. In both studies it was found that changes appear at a very fast rate (the domain “.com” changing daily in some 20% of the sample [12]). A structured study from 2004 discovered instead that the large majority of web-sites sampled do not change the contents of existing pages, but they tend rather to add new content, in doing so copying the content of existing pages [14]. The use of quantitative data to support a “theory” of web evolution has also interested many studies: using graph theory to determine the extension of the web [15]

TABLE II
INFORMATION TECHNOLOGIES SKILLS – SERVER SIDE WEB DEVELOPMENT

Keyword	09/09	10/09	11/09	12/09	01/10	02/10	03/10	04/10	05/10
ASP.NET combinations									
ASP.NET and SQL Server	557	665	580	471	569	634	736	679	651
ASP.NET, SQL Server and IIS	79	68	61	56	68	52	74	85	68
ASP.NET and Oracle	59	65	67	45	55	65	63	49	53
ASP.NET and MySQL	67	45	44	33	32	50	58	27	36
PHP combinations									
PHP and MySQL	387	304	282	241	344	442	444	345	327
PHP, MySQL and Apache	93	86	81	68	84	131	120	90	83
LAMP (Linux, Apache, PHP, MySQL)	173	188	155	145	205	223	215	163	173
WAMP (Windows, Apache, PHP, MySQL)	15	12	12	8	16	23	28	6	12
PHP and SQL Server	65	61	53	36	42	49	83	58	42
PHP and Oracle	29	31	29	34	32	45	47	28	39
PHP, MySQL and Oracle	13	14	20	21	24	24	38	14	19
J2EE combinations									
J2EE and Oracle	148	112	98	124	137	160	150	157	187
J2EE and SQL Server	51	38	27	29	39	42	61	40	45
J2EE Tomcat and MySQL	11	14	14	6	17	21	15	18	37
J2EE Tomcat and SQL Server	6	7	2	2	8	10	14	5	7

TABLE III
INFORMATION TECHNOLOGIES SKILLS – OPERATING SYSTEMS

Keyword	09/09	10/09	11/09	12/09	01/10	02/10	03/10	04/10	05/10
Microsoft Operating System									
Win XP	146	143	130	118	223	164	149	171	146
Win Server 2003	83	73	95	58	102	81	104	112	84
Win Server 2008	11	12	27	24	19	9	24	24	12
Windows Vista	106	94	81	64	104	90	89	104	112
Windows 7	10	18	18	13	40	40	53	50	70
Generic Windows	1,416	1,191	1,284	992	1,424	1,385	1,293	1,295	1,248
Linux Operating System									
RHCE	15	21	7	10	19	8	15	9	12
Ubuntu	8	8	8	16	15	11	19	14	16
SuSE	17	18	20	17	14	16	10	13	19
OpenSuSE		1							
Debian	13	25	23	18	19	17	16	18	26
Unspecified Linux	786	667	746	602	828	902	935	829	805
Unix and Unix-like									
Solaris	183	150	164	110	146	211	202	119	145
IRIX	7	1	4	1	1		1		
AIX	66	62	73	54	74	79	69	54	75
OpenSolaris						4	2		
OpenBSD		3	2		3				
FreeBSD	8	2		3	4	5	6	6	
NetBSD				1					
Unspecified Unix	772	607	730	522	759	803	720	632	663
Macintosh									
MacOSx	30	19	27	16	24	27	40	34	44

TABLE IV
COURSES AND MODULES FROM THE SELECTED UNIVERSITIES' WEB SITES

ID	Degrees	Modules	Keywords found (out of 30)	Density (all keywords)
1	47	217	21	7.260e – 03
2	35	n/a	4	6.798e – 03
3	21	162	22	2.093e – 03
4	20	58	1	9.851e – 05
5	20	n/a	10	9.175e – 03
6	16	n/a	2	2.777e – 04
7	15	104	3	1.025e – 04
8	13	n/a	4	6.147e – 04
9	9	70	12	1.641e – 03
10	6	n/a	1	3.500e – 04

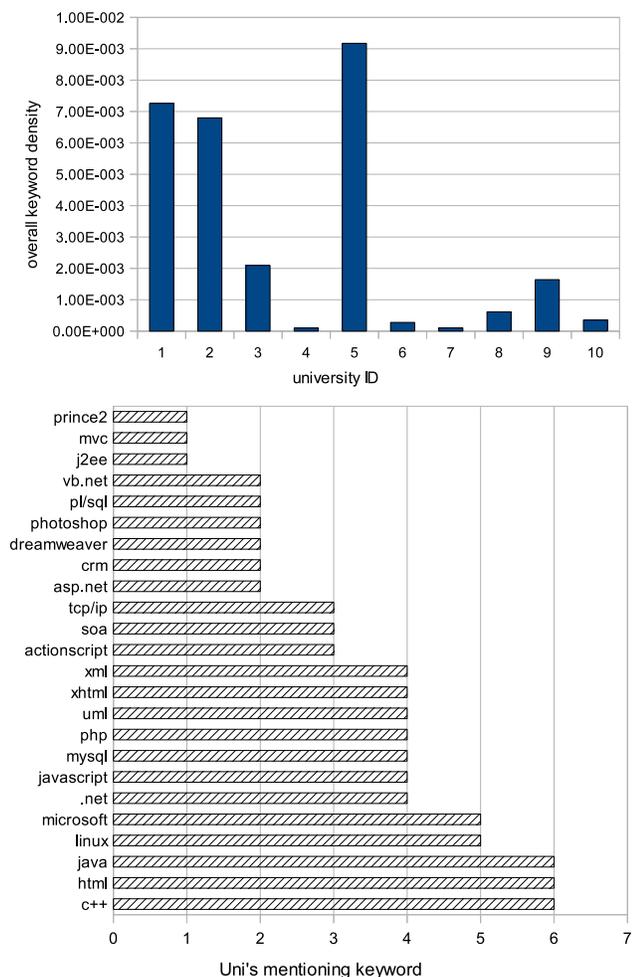


Fig. 3. Most cited keywords by number of UK universities (below) and keyword density in courses and modules descriptions (top)

including attempts to recalibrate web-crawlers to detect future trends of the evolution of the WWW at large [16]. Albeit an external limitation, this study instead focuses on the evolution of contents of just one specific web-site (monster.com): this was done, first and foremost, because a “generic” study of how websites change is outside the scope of this work; secondly, the objective of this paper was to try and to produce a rationale of cause-effect in the co-evolution of websites.

On the whole, there has been very little research on the recent trend to utilize fully the Web as part of the recruiting process, both from the provisions and/or the demand of jobs. Little empirical evidence is available about the actual effects of Internet job searches. One of the most cited studies on the topic deals with the parsing of large jobs websites, in order to extract keywords used in the searches by job hunters [17]. A study regarding the re-employability of the on-line job searchers found instead that unemployed Internet job searchers do not become reemployed more quickly than “traditional” job hunters [18]. Some attention has also been given to the “on-line recruiting service” as one of the most important e-

businesses of the last few years, and proposals for gathering information about users behavior and build an adaptive job recommender application [19].

On the other hand, the description of jobs by job recruiters (specifically on dedicated portals, like monster.com) was studied in [20]. The study focuses more specifically on the ways the recruiters use for advertising new available positions: the clear indication is that the description of jobs on monster.com tend to focus more on the firm attributes, and only secondarily on employee advancement. Other studies focused on the monster.com portal focused on the demand for IT professionals, and argued that course deliverers should keep an eye on the outcomes of such studies: the design and development of new curricula to prepare students for the job market should be based on actual skills requested by the industries [5], [10]. As far as we know, this is the first study assessing the on-line requests for IT skills in the UK, and trying to match these requests to the provision of UK academic institutions.

VI. THREATS TO VALIDITY

Like any other empirical study, the validity of ours is subject to several threats. In the following, threats to internal validity (whether confounding factors can influence your findings), external validity (whether results can be generalized), and construct validity (relationship between theory and observation) are illustrated.

A. Internal Validity

The following threats to internal validity have been detected:

- on the “demand” side, this research has not considered printed job posts. An in-depth analysis of the job adverts in the press would have required a level of manpower that we were not able to commit at this stage. While many prospective workers do not use the Internet for job hunting, it was thought that IT workers and recruiters were both most likely to use the Internet.
- The academic institutions were ranked based on the number of their students. This criterion was used because it is argued that, when serving more students, the amount and relevance of available information relative to the provided courses should increase.
- The descriptions of courses and modules considered in this study were only at the undergraduate level: first, because one can argue that students do not wish to enrol in a BSc and an MSc before applying for a job; and second, because not all the universities in the sample offer a similar range of BSc and MSc courses.

B. External Validity

Some threats to external validity have also been detected:

- Only the **monster.com** website has been considered for study, but not other jobs on-line. Given the fact that Monster greatly outnumbers the other advertisers, and considering the great effort needed for the development of the software used for the data collection, it was decided

that at this stage only data from Monster UK would have been collected and analysed. While it is a limitation, this nevertheless provides a good starting point for our analysis.

- On the “offer” side, only 10 UK Universities have been considered for study: given the difficulty to parse diverse information, it was thought that 10 out of 80 could represent an initially adequate sample.
- Having used only on-line jobs poses also a threat to external validity: while jobs might be both in the printed press and on-line, a number of job posts will be listed in the printed press only. They are not suitable for any kind of quantitative analysis as the one that has been carried for this paper with the resources available to the authors.

C. Construct Validity

Further threats to construct validity are:

- We assumed that the description of modules and courses is the only interface that prospective students use when looking for information about a course or module. We also assumed that the course and module descriptions are accurately prepared by marketing teams to actively reflect all the information needed by prospective students to make an informed choice.
- We also assumed that IT prospective employees use mostly the web for looking for jobs. Government statistics suggest that 76% of the UK adult population has regular access to Internet and 70% of the population does have Internet access from home [21]. We are aware that this approach can be used to monitor on-line job listings only, and is limited to analysing the opportunities of job seekers with Internet access. That said, it is a reasonable assumption to expect that prospective workers in IT have some level of Internet access and use Internet for job seeking. In future, further research might provide more detailed information on this aspect.

VII. CONCLUSIONS AND FUTURE WORKS

The types and forms of new IT skills needed by companies to stay competitive in the global market act as powerful drivers for the evolution of contents of various on-line actors: jobs agencies, higher education institutions and standardisation bodies. While it has been reported that IT course providers have often responded well to the needs of industry, by massively reshaping the computer science curricula in the last 10 years, this paper has tried to view the problem from the point of view of a prospective higher education IT student, accessing only on-line resources of both demand of IT skills (the monster.com website), and offer of such skills (the websites of UK universities).

The results are both interesting and depressing: the monster.com IT section clearly responds to the needs of companies, by promptly updating the job listings relative to specific skills. Instead of an agreed terminology proposed by a standardisation body, IT companies tend to use specific keywords strictly related to IT skills. Together with established platforms

(Microsoft, Unix and Macintosh) a new surge of Open Source software skill-sets (Linux in particular) is visible and sustained throughout the parsing of the last 9 months.

On the other hand, the provision of IT skills suffers at least from two macroscopic problems: the first is that, in a sample of 10 UK universities, half of the analysed courses do not offer enough information of what a prospective student will learn in the course of the selected BSc. The second is that, based on the descriptions, what the courses offer is not aligned with what is demanded by IT companies, and mirrored by on-line job agencies. Only very two universities seem to be able to use appropriate keywords in the on-line descriptions of what students should expect.

While it is always challenging to keep an eye on the various needs of the market, and to offer an adequate provision of IT courses, we hope that the what emerges in the most dynamic and prompt UK universities can be used as an example for others in the description of their IT courses. This does not need a mass reshaping of the courses, but a better focus on what are their contents and what skills will be provided to students.

In terms of future works, a number of possible research directions are left open:

- Monitor what type of jobs are more required by UK companies, and how it changes in time
- Monitoring what type of academic or professional qualifications if any at all are more required by UK companies
- Monitoring the ups and downs of the market
- Extending the research to more areas of IT
- Replicating the research to areas other than IT
- Extending the number of sources monitored (e.g. monitoring other web sites that do include job postings, as the Guardians one)
- Ex-post interviews with companies that have posted adverts on-line, to analyse whether they are satisfied or not.

REFERENCES

- [1] Office for National Statistics, “Labour market statistics – statistical bulletin (november 2009),” <http://www.statistics.gov.uk/pdfdir/lmsuk1109.pdf>, 2009.
- [2] SFIA, “Framework reference,” <http://www.sfia.org.uk/cdv3/ai1005484.html>, 2009.
- [3] UCAS, “Provisional final figures for 2009,” http://www.ucas.com/about_us/media_enquiries/media_releases/2009/2009-10-21, 2009.
- [4] S. Gambill and J. Maier, “CIS/MIS curriculums in AACSB and non-AACSB accredited colleges of business,” *Journal of Information Systems Education*, vol. 12, no. 1-2, pp. 59–67, 1998.
- [5] K. Koong, L. Liu, and X. Liu, “A study of the demand for information technology professionals in selected Internet job portals,” *Journal of Information Systems Education*, vol. 13, no. 1, pp. 21–28, 2002.
- [6] X. Liu, L. C. Liu, K. S. Koong, and J. Lu, “An examination of job skills posted on internet databases: Implications for information systems degree programs,” *Journal of Education for Business*, vol. 78, no. 4, pp. 92–96, 2003.
- [7] V. R. Basili, G. Caldiera, and D. H. Rombach, “The goal question metric approach,” in *Encyclopedia of Software Engineering*. John Wiley & Sons, 1994, pp. 528–532, see also <http://sdqweb.ipd.uka.de/wiki/GQM>.
- [8] comScore, “Job search ranks as fastest growing u.s. online category in 2008,” http://comscore.com/Press_Events/Press_Releases/2009/1/Job_Search_Fastest_Growing/%28language%29/eng-US, 2008.

- [9] C. Girardi, F. Ricca, and P. Tonella, "Web crawlers compared," *International Journal of Web Information Systems*, vol. 2, no. 2, pp. 85–94, 2006.
- [10] B. Prabhakar, C. R. Litecky, and K. Arnett, "It skills in a tough job market," *Commun. ACM*, vol. 48, no. 10, pp. 91–94, 2005.
- [11] Netcraft, "November 2009 web server survey," http://news.netcraft.com/archives/web_server_survey.html, 2009.
- [12] J. Cho and H. Garcia-Molina, "The evolution of the web and implications for an incremental crawler," in *VLDB '00: Proceedings of the 26th International Conference on Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 200–209.
- [13] D. Fetterly, M. Manasse, M. Najork, and J. L. Wiener, "A large-scale study of the evolution of web pages," *Softw. Pract. Exper.*, vol. 34, no. 2, pp. 213–237, 2004.
- [14] A. Ntoulas, J. Cho, and C. Olston, "What's new on the web?: the evolution of the web from a search engine perspective," in *WWW '04: Proceedings of the 13th international conference on World Wide Web*. New York, NY, USA: ACM, 2004, pp. 1–12.
- [15] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," *Computer networks*, vol. 33, no. 1-6, pp. 309–320, 2000.
- [16] M. Toyoda and M. Kitsuregawa, "What's really new on the web?: identifying new pages from a series of unstable web snapshots," in *WWW '06: Proceedings of the 15th international conference on World Wide Web*. New York, NY, USA: ACM, 2006, pp. 233–241.
- [17] B. J. Jansen, K. J. Jansen, and A. Spink, "Using the web to look for work: Implications for online job seeking and recruiting," *Internet Research*, vol. 15, no. 1, pp. 49–66, 2005.
- [18] P. Kuhn and M. Skuterud, "Internet job search and unemployment durations," *The American Economic Review*, vol. 94, no. 1, pp. 218–232, 2004. [Online]. Available: <http://www.jstor.org/stable/3592776>
- [19] D. Lee and P. Brusilovsky, "Fighting Information Overflow with Personalized Comprehensive Information Access: A Proactive Job Recommender," in *Proceedings of the Third International Conference on Autonomic and Autonomous Systems*. IEEE Computer Society, 2007, p. 21.
- [20] K. Backhaus, "An exploration of corporate recruitment descriptions on Monster.com," *Journal of Business Communication*, vol. 41, no. 2, p. 115, 2004.
- [21] Office for National Statistics, "(2009) internet access," <http://www.statistics.gov.uk/cci/nugget.asp?ID=8>, 2009.