Teknillinen korkeakoulu Tietotekniikan osasto
Tietojenkäsittelyopin laboratorio B
Helsinki University of Technology Department of Computer Science and Engineering
Laboratory of Information Processing Science B
Espoo 2005 TKK-TKO-B156

TREND ANALYSIS OF JOB ADVERTISEMENTS: WHAT TECHNICAL SKILLS DO SOFTWARE DEVELOPERS NEED?

Sami Surakka

EPS EPS

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY

EPS EPS

Helsinki University of Technology
Laboratory of Information Processing Science
P.O. Box 5400
FIN – 02015 HUT
URL: http://www.cs.hut.fi/english.html

ISBN 951-22-7579-1 (PDF)
ISSN 1239-6893
URL: http://www.cs.hut.fi/Publications/Reports/B156.pdf

This publication is available only as a PDF file. It is purposed to be printed on A4 paper.

HELSINKI UNIVERSITY OF TECHNOLOGY http://www.hut.fi/	ABSTRACT PAGE
Department/laboratory and URL/Internet address Department of Computer Science and Engineering Laboratory of Information Processing Science http://www.cs.hut.fi	Publisher Helsinki University of Technology Department of Computer Science and Engineering Laboratory of Information Processing Science

Author
Sami Surakka

Title
Trend analysis of job advertisements: What technical skills do software developers need?

Abstract

An American magazine Computerworld was used to find the most common technical skills sought in job advertisements for software developer positions. Data were collected from every second year from 1990 to 2004 (*N*=1291, *n*=112...265). During this period, the mean of the number of required technical skills increased from 3.6 to 7.7.

Unlike in previous longitudinal researches, also distributed technology skills were analyzed thoroughly because as a consequence of World Wide Web technology, these skills are required now more often than ten years ago. Skills were divided into five categories: programming language skills, operating systems skills, database skills, networking skills, and distributed technology skills. In these five categories, the increase in distributed technology skills was very strong: the proportion increased from 0% in 1990 to 65.1% in 2004. The proportions increased strongly also in the other categories than networking skills. Thus, the technical requirements have changed as more versatile. In addition, the results about individual skills such as C++, Oracle, Java Server Pages, and TCP/IP are reported.

Keywords and classification
job advertisement analysis, need assessment, skills, software developer, trend analysis

Place	Year	Number of pages	Language of publication	Language of abstract
Espoo	2005	30	English	English

ISBN (printed)	ISSN and number or report code (printed)
ISBN (electronic)	**ISSN and number or report code (electronic)**
951-22-7579-1 (PDF)	1239-6893, TKK-TKO-B156

URL (Internet address)
http://www.cs.hut.fi/Publications/Reports/B156.pdf

Supplementary bibliographic data (edition, figures, tables, appendices)
Appendices: 2 (3 pages).

Table of contents

1 Introduction

In the present research, the goal has been to identify technical skills that are important for software developers' work. The research origins from the need to better understand what topics and skills should be included in the master's level education of software systems specialization in a technical university. Job advertisements were analyzed to find out the most common technical skills sought in various software developer positions. The present research is a part of my on-going doctoral thesis work where different research methods have been used to evaluate software systems education (e.g., job advertisement analysis as in the present research, the Delphi method targeted at software developers, and questionnaires targeted at graduating master's students).

Unlike the most previous job advertisement analyses, the present research was targeted particularly to technical skills needed in software developer positions. Two main questions were: (1) Has the number of required technical skills increased during the past 15 years? Todd, McKeen, and Gallupe (1995) reported that the number of technical phrases in job advertisements for programmer positions increased from the mean of 2.2 in 1970 to 4.2 in 1990. Has this increase continued after the year 1990? (2) In particular, how has the number of required distributed technology skills increased? World Wide Web technology was released in 1993. After this, the number of web sites has increased rapidly. As a consequence, skills related to distributed systems are now required more often than ten years ago.

During my previous job advertisement analysis (Surakka 2004), I noticed that the answer to the first question is likely to be positive. Anyhow, in the present research the goal was to get a more detailed view to these changes using a trend analysis when the samples were collected from every second year during the period of 15 years.

The results of the present study might be useful for the training departments of companies, training institutes, and curriculum designers in universities. Students might use the results when they are selecting elective computer science courses, especially in industry-oriented master's programs. A software developer working in industry might want to compare his or her skills to the results. The results might even reveal something about which technologies were used in new projects if it is assumed that new employees are often hired for such projects. From this viewpoint, the competition between Microsoft's and Sun's technologies for distributed systems is particularly interesting (Section 4.4.5).

The structure of the report is the following. First, the related work is considered in Section 2. In Section 3, the research method is described. The results are presented and analyzed in Section 4. Finally, the research is evaluated, conclusions are drawn, implications to education are considered, and some possibilities for future research are presented.

2 Related work

For any reader wishing to get an overview of research into IT needs assessments, two good starting points are the papers by Nakayama and Sutcliffe (2000; 2001). Most previous relevant research has been carried out by educators and researchers who work for information systems (IS) or information technology (IT) degree programs, not for computer science (CS) or software engineering (SE) programs. 37 publications that are related to the present research were found. 32 (86%) of them were from the field of IS/IT. The results have been typically published in publications like the MIS Quarterly, Journal of Computer Information Systems, and the proceedings of ACM's Special Interest Group for Computer Personnel Research (SIGCPR, currently merged with SIGMIS). These 37 relevant publications were classified according to the research methods: 17 (46%) were content analyses of job advertisements, 9 (24%) were surveys, 4 (11%) were literature surveys or research in progress reports, 1 (3%) was an interview, and 6 (16%) used more than one method.

From these 17 job advertisement analyses mentioned above, seven are longitudinal researches as the present research (Athey & Plotnicki 1998; Gallivan, Truex, & Kvasny, 2002; Litecky, Prabhakar, & Arnett 1996; Maier, Clark, & Remington, 1998; Prabhakar, Litecky, & Arnett, 1995; Todd et al., 1995; Trower 1995). However, Litecky et al. (1996) and Prabhakar et al. (1995) contain results only from two or three subsequent years from the period 1993–1995. In other longitudinal researches the periods were 6–21 years. The periods used by Litecky et al. and Prabhakar et al. are so short that I have classified their researches as cross-sectional researches. The five other longitudinal researches are the most relevant to the present research. Later in Sections 3 and 4, the comparisons of methodological choices and results are targeted to these.

All these five previous longitudinal researches were targeted at IS/IT field. Next, these researches are shortly presented:

- Athey and Plotnicki (1998) analyzed individual technical skills such as Cobol, Oracle, and Windows, and categories such as mainframes and minicomputers. They reported results from ten different cities as well. The samples came from the years 1989, 1992, 1993, and 1996.
- Gallivan et al. (2002) counted, for example, proportions of different job titles and averages of skills per advertisement. The samples were from the years 1988, 1995, and 2001. The coding scheme (p. 13) included also categories for individual technical skills such as Cobol and C but they did not report these results.
- Maier et al. (1998) analyzed mostly individual technical skills such as C, Cobol, and Unix. The samples were from the years 1978/79, 1983/84, 1988/89, and 1993/1994.
- Todd et al. (1995) analyzed technical, business, and systems skills for programmers, systems analysts, and managers. The samples were collected from the years 1970, 1975, 1980, 1985, and 1990. From these five researches, Todd et al. is the most relevant to the present research because it had also detailed results about programmers.
- Trower (1995) analyzed individual technical skills and combined them as categories such as "3GL" (e.g., Cobol, Fortran, or Pascal). The samples were collected from the period of 1990–1995 at one-year intervals. His research was particularly interesting from the viewpoint of the present report because he reported also results about distributed technology skills.

Beside the previous scientific publications, there is one non-scientific report series that is worth mentioning. A British company Salary Services Ltd. has conducted job advertisement analyses for the past 15 years. The data have been collected from several British newspapers and web recruiting services. In particular, this report series is interesting because it is problematic to get data from web recruiting services (see Sections 3.1 and 5.3). At least in a recent report (Salary Services 2004a), the data were processed automatically and the sample was much greater than in the previously mentioned scientific longitudinal researches. The report ranks 150 individual technical skills from the years 1999–2003 with an one-year interval. Although this report is not a scientific publication, in my opinion, it is convincing. Unfortunately, the report is protected by copyright that makes comparisons of results more difficult (see Sections 4.4.2–4 and Appendix B).

The following characteristics are different in the present research when compared with the most previous job advertisement analyses: (a) The present research is limited to software developer positions and is more detailed in the area of technical skills than previously. (b) Advertisements for more scientific or engineering-oriented positions have sometimes been excluded from the previous samples for an obvious reason: these positions are not very suitable for IS/IT graduates. I work at a university of technology and, therefore, included also more scientific or engineering-oriented positions in the sample. (c) When compared with previous longitudinal researches, there are several differences in methodology. These differences are explained next in Section 3.

3 Method

Content analysis is a method that is widely used in communications research. Some good properties of content analysis are: (a) it is a non-disturbing method because data occur regardless of whether the research is carried out or not, and (b) it is often possible to get a representative sample. In the present research, a quantitative content analysis of job advertisements was carried out; that is, the frequencies of different phrases such as C++, Oracle, and WebLogic were simply counted.

In the following subsections, the details of the methodology used in the present research are presented. The choices are also compared against the five previous longitudinal researches that are the most relevant to the present research.

3.1 Choice of data source

In the previous longitudinal researches, newspapers or professional magazines have been used as data sources; that is, web recruiting services have not been used. The number of newspapers and magazines has varied from two to ten. The justification for several data sources has been to reduce the effect of possible regional differences. The use of newspapers has been more common than the use of professional magazines. Gallivan et al. (2002) is the only previous research that has used a professional magazine (Computerworld) as a data source. The commercial analysis (Salary Services 2004a, p. 286) used three professional magazines, six newspapers, and six web recruiting services.

In the present research, web recruiting services were not selected because they have not operated long enough for the purpose of the present research and data from the past few years are not publicly available. This was a very problematic situation because according to Salary Services (2004a, p. 2), the web services have dominated the IT job advertising market at least since the year 2000 and therefore, it was not certain if newspapers and professional magazines were still representative data sources. However, in my opinion, (a) it was better to conduct a trend analysis using a magazine or a newspaper than not to conduct a trend analysis at all because data from web services were not available, and (b) the selected data source was the most suitable that was available.

Only one data source was used to keep the amount of work as reasonable. Newspapers such as New York Times were not selected because they were too regional. From various magazines targeted to IT professionals, Computerworld was chosen. Other possible professional magazines would be, for example, Communications of the ACM, IEEE Computer, and IEEE Software. The circulations of these four magazines are approximately as follows (Ulrich's periodical directory, 2004): Computerworld 250,000, IEEE Computer 97,000, Communications of the ACM 85,000, and IEEE Software 23,000. Computerworld is published weekly and the other three magazines six or 12 times per year. The main reasons to choose Computerworld were that it is a national magazine and might be more compelling to advertisers because it has the biggest circulation and is published weekly.

3.2 Sampling

In the previous longitudinal analyses, the periods have varied from six to 20 years and the intervals from one to seven years. Athey and Plotnicki (1998) and Gallivan et al. (2001) used unequal intervals whereas others used equal intervals. In the present study, the year 1990 was chosen as the starting year because WWW technology was released in 1993 and I wanted to get some results also before that year. Every second year was chosen as the interval to get more detailed results from possible trends. Every fifth year was not used because I assumed that changes in distributed technologies, in particular, might be so fast and large during the period that it would be beneficial to use a shorter interval.

Gallivan et al. (2001) and Maier et al. (1998) collected data from four numbers per year. Todd et al. (1995) collected data from each month of the year to avoid the possibility of seasonal or cyclical effects on data. However, in these three papers, it was not reported if there were any seasonal or cyclical effects. Athey and Plotnicki (1998), and Trower (1995) collected data from one number per year. In the present research, one number per year was chosen as sampling strategy. This decision was based on an assumption that there are no significant seasonal or cyclical differences. Normally, the sampling was the number 36. However, in 2002 and 2004 the sampling was three numbers (36–38) to get large enough subsamples from every year. The limit was set to at least 100 positions per year.

3.3 Coding

The selected 12 numbers had altogether 1004 job advertisements that were read and coded manually. In the previous longitudinal researches, apparently one advertisement was used as the unit of analysis. It is possible that one advertisement contains several job titles and one job title contains several positions. In Salary Services (2004a, p. 285) and in the present research, one position was used as the unit of analysis.

Gallivan et al. (2002) wrote (p. 5) “In *Computerworld*, we initially treated ads for the U.S. East coast region and non-East regions separately. Subsequent analysis revealed that there was no statistical difference between the distribution of ads by job category or skill, and we subsequently combined the data for these geographic regions.” In the present research, the advertisements were not separated by regions.

An advertisement was included if it contained at least one suitable position such as “Programmer”, “Programmer Analyst”, “Software developer”, or “Software engineer.” For example, the job titles “C++ developer” and “Web developer” were classified as software developer positions and included. A systems analyst position was included if it included also programming tasks. If this was unclear, a systems analyst position was excluded. A systems programmer position was included if it was rather a developer position than a systems administrator position. A systems programmer position was excluded if this was unclear. The following positions, for example, were excluded: “Business analyst”, “Consultant”, “Database administrator”, “Project manager”, “Quality assurance engineer”, and “Systems administrator.” A position was excluded if no position title was given or it was too general, for example “IT professional.”

In addition, also programmer etc. positions were excluded if the required degree was any doctoral degree or the field of study was not suitable for a computer science graduate, for example, Masters in Electrical Engineering was required. In other words, the position should be suitable for a

Bachelor's or Master's degree graduate from a computer science program. Contractor positions were included and part-time positions excluded.

Each suitable job advertisement described one or more open positions. If the number of positions was not given, it was coded as one position. If the number of positions was not given but text indicated a plural (e.g., "Programmers"), it was coded as two positions. The exact number of positions was used if it was given (e.g., "5 programmers"). Altogether these job advertisements contained 1291 suitable positions, which is the sample of the present research (i.e., N=1291). The number of suitable positions for each year varied from 112 to 265. These are the subsamples of the present research (i.e., n=112...265). The numbers of included advertisements, excluded advertisements, and excluded positions were not counted.

From advertisements of these suitable positions, it was searched for technical skills such as Cobol, Java, SQL, and Windows. These phrases were typically names or abbreviations of different programming languages, operating systems, database vendors, and protocols. Only required skills were included, and desired skills were excluded. A skill was coded as required if it was not mentioned if it was required or only desired.

3.4 Statistical methods

For each individual skill, a proportion was calculated for each year. For example, the sample of the year 1990 had 189 positions and in 77 of these positions, Cobol was required. Thus, the proportion of Cobol was 40.7%. In addition to these proportions for individual skills, several other figures were calculated. These coding principles are explained later before the corresponding results.

The Student t test and the Smith-Satterthwaite procedure (e.g., Milton & Arnold 2003, pp. 347–349) were used to test if the difference between two means was statistically significant. The Z test for proportions (e.g., ibid., p. 324) was used to test if the difference between two proportions was statistically significant. For significance, the following limits were used: not significant $p \geq 0.05$, almost significant $p<0.05$, significant $p<0.01$, and very significant $p<0.001$.

In the previous longitudinal analyses, the use of statistical tests to analyze the results has been rare or non-existing, or at least it is not reported. Gallivan et al. (2001) have apparently used statistical methods but from other analyses I found nothing that would indicate the use of statistical tests.

4 Results

The results are divided into subsections. First, in Sections 4.1, 4.2, and 4.3, some results are presented to get an overview how the requirements have changed. Second, in Section 4.4, more detailed results are presented for individual platform skills, programming language skills, database skills, networking skills, and distributed technology skills. Finally, in Section 4.5 some miscellaneous results are presented.

Below each figure the results of the present research are compared against the results of the previous longitudinal researches and the commercial report (Salary Services 2004a), if possible. The sizes of subsamples and proportions as numbers are presented as tables in Appendix A.

4.1 Proportions of job titles

Figure 1 presents the proportions of four job titles that were used in the present research: programmer, software developer, software engineer, and systems analyst. For example, in 1990 the size of the subsample was 189 positions and the number of programmers was 161. Thus, the proportion of programmers was 85.2%. It can be noticed that the proportion of programmers has decreased and the proportion of software developers and engineers have increased. The proportion of systems analysts has been small for the whole period. Based on the Z test for proportions, the differences between the years 1990 and 2004 are statistically very significant ($p<0.001$) for programmers, software developers, and software engineers, and not significant ($p\geq0.05$) for systems analysts.

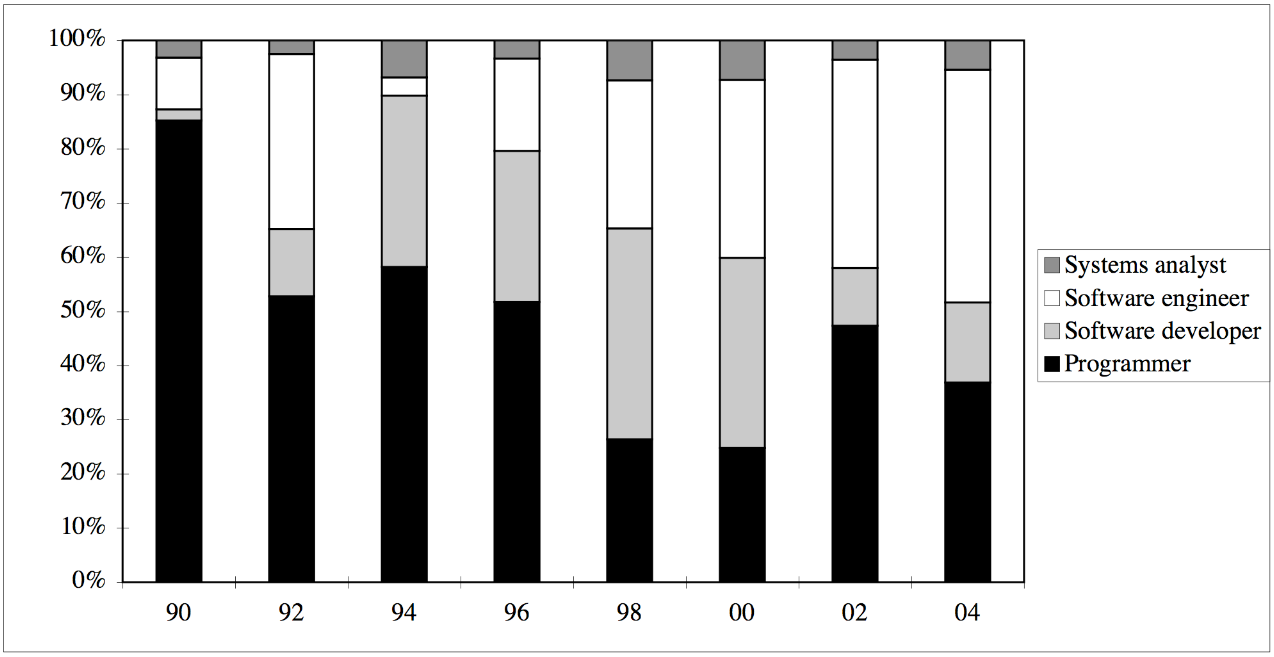

Figure 1. Proportions (%) of four job titles in years 1990–2004.

From previous publications, Gallivan et al. (2002) is the most suitable for comparison. They (p. 12) reported that the proportion of programmers decreased from 64% in 1988 to 20.5% in 2001. During the same period, the proportion of software engineers increased from 3.5% to 16.1%%. The proportion of systems analyst was quite low: 1.3 in 1988 and 4.2 in 2001. They did not have the category "Software developer" but the proportion of the category "Web developer" was 4.7% in 2001. Apparently, this category was used only in the subsample of the year 2001. Their sample included all kinds of IT positions but nevertheless, the results are similar to the present research: the proportion of programmers has decreased, the proportions of other developer titles have increased, and the proportion of systems analysts has been low during the whole period.

4.2 Number of required technical skills

The number of required individual technical skills was calculated for each position. For example, if Cobol and DB2 were required, the number of skills would be two. The minimum number was used if alternatives were given. For example, for the text "C++ or Java" the number of skills would be one. The mean value of these numbers was calculated for each year. Figure 2 presents these means for three categories: "All", "Programmer", and "Other." The software developer, software engineer, and systems analyst positions were combined as the category "Other" because some of the subsamples were too small to be presented alone. It can be noticed that the means have increased during the period. For example, the mean of the category "All" increased from 3.57 in 1990 to 7.66 in 2004. Based on the Student t test and the Smith-Satterthwaite procedure, the differences between the means of the years 1990 and 2004 are statistically very significant ($p<0.001$) for the categories "All" and "Programmer", and significant ($p<0.01$) for the sample "Other."

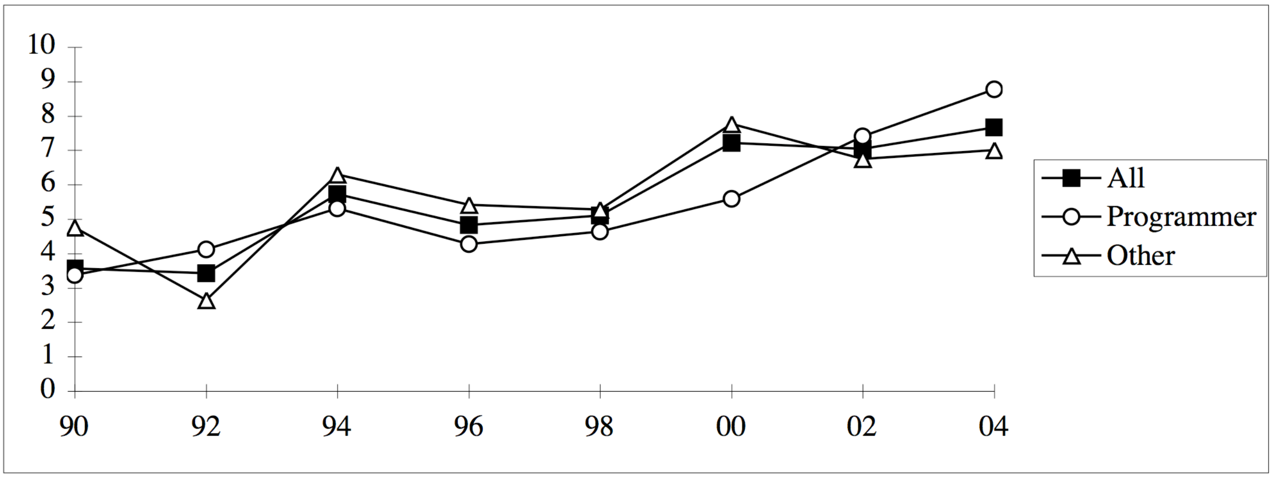

Figure 2. Means of number of required technical skills in years 1990–2004.

As mentioned in Section 1, Todd et al. (1995) reported that the number of technical phrases in job advertisements for programmer positions increased from the mean of 2.2 in 1970 to 4.2 in 1990. They did not report if this difference was statistically significant, and it is not possible to test this because standard deviations were not reported, and the standard deviations or the original data were no longer available.

Maier et al. (1998, p. 38) and Gallivan et al. (2002, p. 7) have reported similar results. However, their samples included all kinds of MIS or IT positions, not just programmers. According to Maier et al., the mean of skills per advertisement was 2.63 for the time period 1978/1979 and 3.50 for 1993/1994. Gallivan et al (2002, p. 7) reported that the mean of technical skills increased from 3.0 skills per advertisement in 1988 to 4.2 in 2001. Maier et al. and Gallivan et al. neither reported if these differences were statistically significant nor standard deviations. I did not ask more information via e-mail from them because the samples included all kinds of positions.

4.3 Proportions of five main skill categories

During this part of analysis, the following criteria were used: (a) at least one common programming language skill (C, C++, Cobol, Java, or Visual Basic), (b) at least one operating systems skill (e.g., AS/400 or Windows NT), (c) at least one database skill (e.g., DB2, Oracle, or SQL), (d) at least one networking or computer networks skill (e.g., LAN or TCP/IP), and (e) at least one distributed technology skill (e.g., client/server or ASP). These five categories are called "Programming language", "Operating systems", "Database", "Networking", and "Distributed technology." For each category proportion was calculated. For example, in the subsample of 1990 were 189 positions and in 92 of these at least one common programming language was mentioned. Thus, the proportion is 48.7%. Similar results for each year and for all five groups are presented in Figure 3.

It can be noticed that the proportions for every category have increased. However, the increase of the category “Networking” has been small and no trend can be noticed. Based on the Z test for proportions, the differences between the proportions of the years 1990 and 2004 are statistically not significant ($p \geq 0.05$) for the category “Networking” and very significant ($p<0.001$) for other categories.

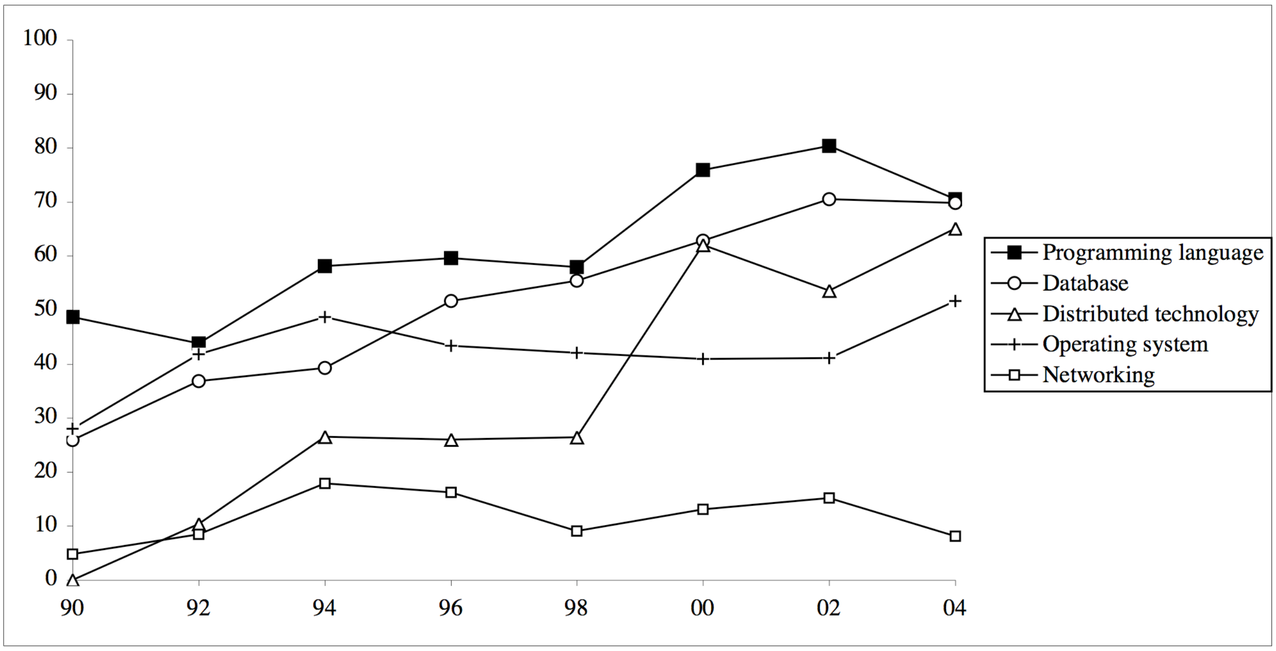

Figure 3. Proportions (%) of at least one skill in five skill categories in years 1990–2004.

Todd et al. (1995, p. 9) used the skill categories “Database” and “Operating systems” that are close enough to the corresponding categories in the present research. They did not report proportions for these categories but they reported the numbers of phrases and the sample sizes. Based on these results, the proportions of the categories “Database” and “Operating systems” have apparently increased during the period 1970–1990. However, it is not possible to calculate proportions in the same way as in the present research. Athey and Plotnicki (1998, p. 76) reported that over 70% of all job opportunities some knowledge of relational database technology. However, they did not report from which year this proportion was; their period was 1989–1996. They reported the increasing trends of individual database skills. Also Trower (1995, p. 599) reported the increasing trend of the category “Relational DB.”

From previous researches, Athey and Plotnicki (1998), and Trower (1995) are the most suitable for comparing the category “Distributed technology.” Trower (p. 599) reported that the number of ads in the category “Client/Server” increased from approximately from 40 in 1990 to 220 in 1995. It is not possible to count the proportions because the sizes of the subsamples were not reported. However, Trower reported an index of all job advertisements as well. The numbers of the category “Client/Server” increased a lot faster than the index did. Anyhow, a more serious problem for comparison is that his coding principles for the client/server skills are problematic. He wrote (p. 598): “Finally, the leading client/server skills mentioned in the 1995 want ads are Windows (88 ads) and OS/2 (37 ads) for the GUI front-ends to client/server applications.” In my opinion, Windows and OS/2 should be classified as operating systems skills, not as a distributed technology skills.[1]

Athey and Plotnicki (1998) reported that the proportion of "Client Server" was 6.9% in 1993 and 7.4% in 1996. The proportion of 1993 matches well enough with the results of the present research but their result for the year 1996 is a lot smaller than the result of the present research. They did not report proportions of "Client Server" for the subsamples of the years 1989 and 1992.

Trower (1995, p. 599) reported that the number of advertisements of the category "Network" increased during the period 1990–95 faster than the index did. This matches with the results of the present research. Other previous researches had no suitable categories or results for comparison. However, Athey and Plotnicki (1998, p. 84) wrote "Certainly networking is becoming more widespread. However, the number of advertised job opportunities in this area are surprisingly small."

I found no previous results that would be suitable for comparing the category "Programming language" because previously, for example, the categories "2GL" and "3GL" have been used.

4.4 Proportions of individual technical skills

Next, more detailed results are presented from individual technical skills. The order of presentation is as follows: platform skills, programming language skills, database skills, networking skills, and distributed technology skills.

4.4.1 Platform skills

In Figure 4, the proportions of different platform skills are presented. The advertisements were classified into the following platform skill categories: Macintosh, "Mainframe or midrange", Unix, "DOS or Windows", and Cross-platform. For example, Windows refers to those positions where some Windows operating system or Windows based software such as Visual Basic or SQL Server was mentioned. Products that are available for both Windows and Macintosh (e.g., Word and Excel) were classified as Windows if Macintosh was not explicitly mentioned. The category "Cross-platform" refers to positions where only cross-platform products such as Oracle were mentioned. If for example Oracle and Visual Basic was mentioned, the position was classified into category "DOS or Windows" only.

It can be noticed that proportion of categories "DOS or Windows" and Unix has increased when the proportion of category "Mainframe or midrange" has decreased. The proportion of Macintosh has been very low during the whole period. Based on the Z test for proportions, the differences between the years 1990 and 2004 are statistically very significant ($p<0.001$) for the categories "DOS or Windows", "Mainframe or midrange" and Unix.

---

1 To be fair to Trower, it would be understandable if he had problems to classify distributed technologies. Even nine years later, I had most problems with this category. For example, I reanalyzed data about distributed technology skills four or five times because I made progressive changes to the classification principles.

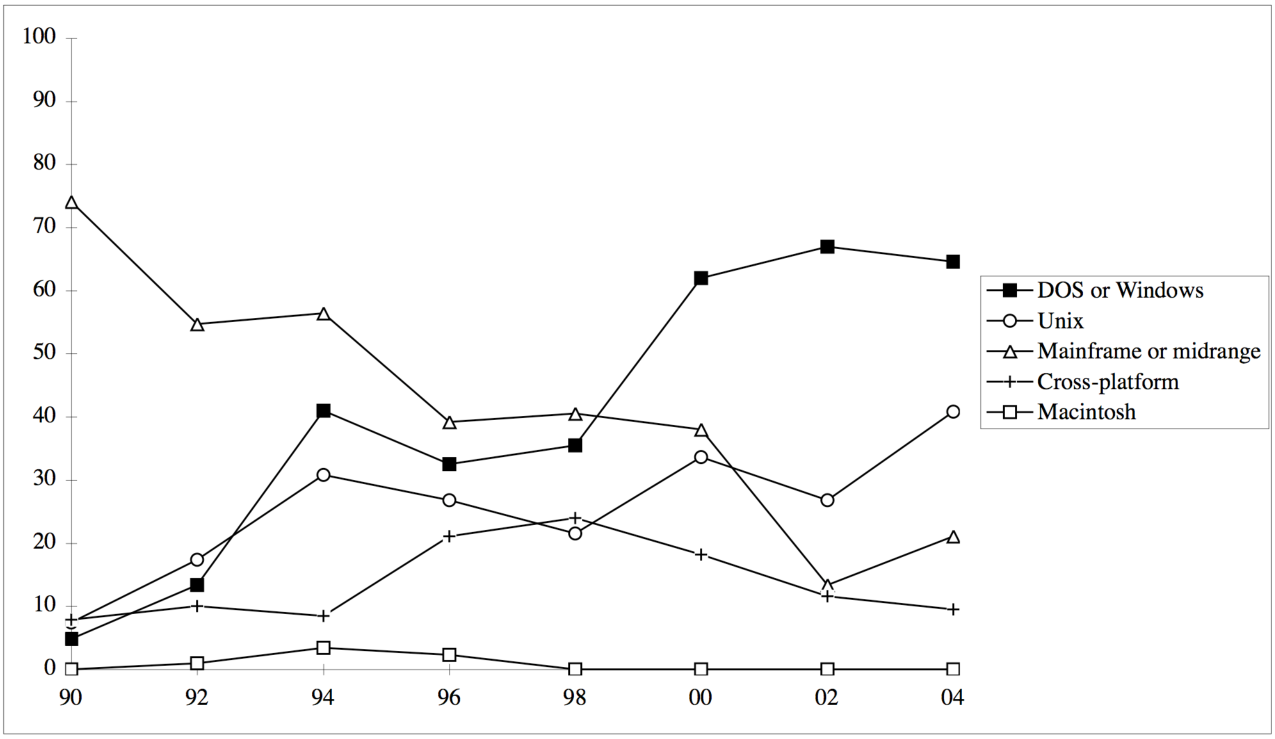

Figure 4. Proportions (%) of different platforms in years 1990–2004.

Maier et al. (1998) reported results about mainframe/midrange, Macintosh, and Unix that are suitable for comparison. Unfortunately, they did not have the category for Windows. The data were collected from four time periods: 1978/1979, 1983/1984, 1988/1989, and 1993/1994. The proportions for Unix increased from 0.2% in 1978/1979 to 25.5% in 1993/1994. The proportions of Macintosh (the category "APPLE") were low in all time periods: the maximum was 3.1% in 1993/1994. Both these values match quite well with the results of the present study. Comparing the category "Mainframe or midrange" against their results was more difficult because the categories were so different. However, for example, the proportions of the category "IBM_AS400" were 29.8% in 1983/1984, 33.6% in 1988/1989, and 17.7% in 1993/1994. Thus, also their results show that the proportion has decreased. In the time period 1978/1979 the category "IBM_AS400" was missing but the proportion of category "IBM" was 25.8%. It appears that the category "IBM_AS400" has been used instead of the category "IBM" after the period 1978/1979.

Todd et al. (1995, p. 8) used the categories "Mainframe", "Mini", Desktop", and "Other." They did not report proportions for these categories but they reported the numbers of phrases and the sample sizes. Based on these results, the proportion of the combined category "Mainframe or Mini" has apparently stayed as approximately the same during the period 1970–1990. However, it is not possible to calculate proportions in the same way as in the present research.

Trower (1995, p. 599) reported that the number of advertisements in the category "Mainframe" decreased during the period 1990–1996. This matches with the results of the present research. He did not report results from other platforms.

Athey and Plotnicki (1998, pp. 79 and 82) reported the following changes between the years 1989 and 1996: "Mainframes" from 17.4% to 5.1%, "Minicomputers" from 35.1% to 15.3% in 1996, "Windows" from 1.7% to 35.1%, and "Macintosh" from 1.6% to 5.2%. Their results of "Mainframes", "Minicomputers", and "Windows" are lower than in the present research but the trends are similar. Their results for "Macintosh" are somewhat greater than in the present research.

In their results, the greatest proportion 7.6% was from the year 1993 which matches reasonable well with the shape of the curve in Figure 4.

4.4.2 Programming language skills

In Figure 5, the most common programming language skills are shown. Only five programming language skills were selected into the figure to keep it readable. The selected languages are such that their proportions have been at least 10% in some year during the period 1990–2004. There are also other such languages but these five were selected into the figure because most of them are probably used in education. More detailed results can be found in Appendix A where results from ten programming languages are presented.

It might help to interpret the figure if one considers that Java was released in 1995, Visual Basic in 1991, but C++ already in 1985, and C and Cobol even earlier than C++. It can be noticed that in the year 2004, Java was the most commonly required programming language. In addition, the proportion of Java has increased strongly. Also proportions of C++ and Visual Basic have increased strongly but this has taken a longer period than for Java. As one might expect, the proportion of Cobol has decreased strongly: from approximately 40% in 1990 to 3% in 2004. Based on the Z test for proportions, the differences between the years 1990 and 2004 are statistically very significant ($p<0.001$) for C++, Cobol, Java, and Visual Basic.

The proportion of C has been quite constant and it is not possible to notice any trend. The proportions have varied, roughly, between 15–30% during this period. However, the proportion in 1990 is considerably smaller than in 2004 (15.3% and 29.5%, respectively). Based on the Z test for proportions, this difference between the years 1990 and 2004 is statistically significant ($p<0.01$).

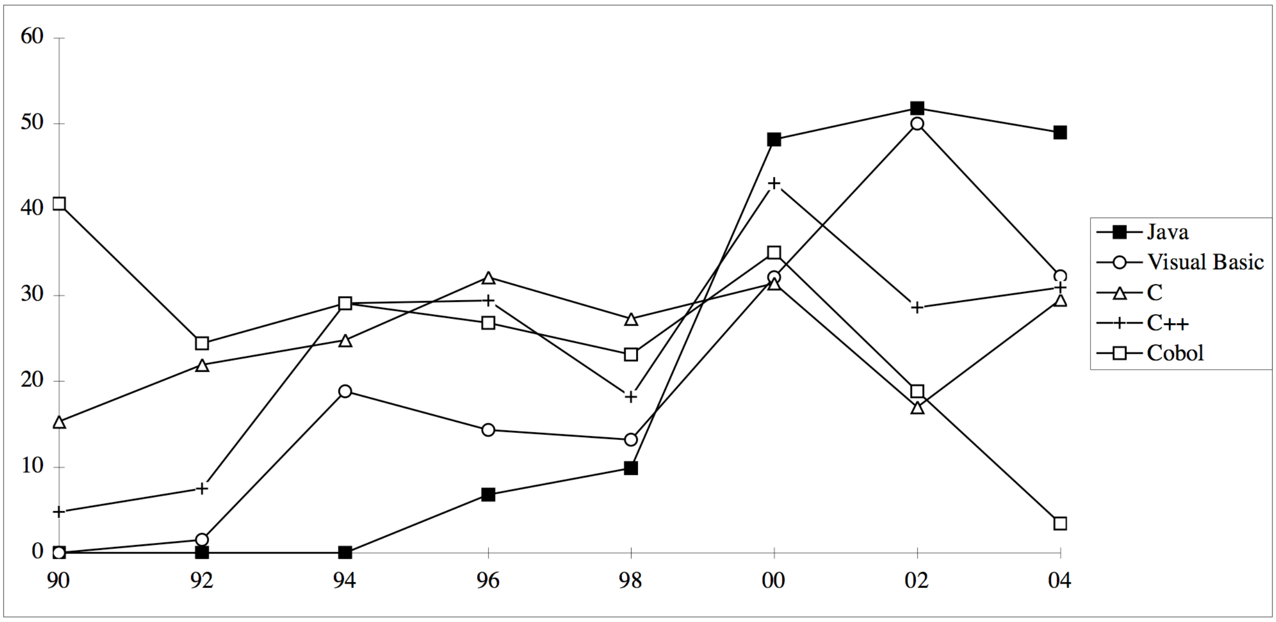

Figure 5. Proportions (%) of different programming language skills in years 1990–2004.

Trower (1995) used categories “3GL” and “4GL” that are too general for comparison. Results by Todd et al. (1995) and Maier et al. (1998) are suitable for comparing Cobol, C, and C++. Todd et al. (1995, p. 9) used the categories “2GL”, “3GL”, “4GL”, and “COBOL.” From these four categories, “2GL”, “3GL”, and “4GL” are too different from the categories of the present research for comparison. They did not report the proportions of Cobol but I calculated the proportions from the other results for programmers. The proportions of Cobol were 41.7% in 1970, 52.5% in 1975, 46.9% in 1980, 47.4% in 1985, and 45.0% in 1990. Based on the Z test for proportions, this difference between the proportions of 1970 and 1990 is statistically not significant ($p \geq 0.05$). The proportion of 1990 in their research (45.0%, n=171) is somewhat greater than in the present research (40.7%, n=189). Based on the Z test for proportions, this difference is statistically not significant ($p \geq 0.05$). Thus, their results of Cobol match quite well with the results of the present research.

Maier et al. (1998, p. 41) reported the proportions for Cobol. In addition, the item “C, C++” was presented in the category listing but in the result table the item was only “C.” Thus, it is unclear if this means only C or “C or C++” but apparently it means “C or C++.” In Table 1 are presented their sample sizes and results for these two categories. It can be noticed that the proportions of Cobol have been decreased and the proportions of “C or C++” have increased. Based on the Z test for proportions, the differences of these two categories between the years 1978/79 and 1993/94 are statistically very significant ($p<0.001$). Thus, their results show similar trends as the present research. Their proportions of Cobol in the years 1988/89 and 1993/1994 are smaller than in the present research in the year 1990 but this is reasonable because their sample included IT positions of all kind, not just software developers.

Table 1. Proportions of Cobol and “C or C++” from 1978/79 to 1993/94.
Source: Maier et al. (1998, pp. 38 and 41).

Years	**Sample size**	**Cobol (%)**	**C or C++ (%)**
1978/79	1795	50.2	0.2
1983/84	1632	40.0	6.5
1988/89	1997	35.9	13.9
1993/94	1886	23.3	29.3

The results by Athey and Plotnicki (1998, p. 75) that are suitable for comparison are presented in Table 2 where the columns are ordered according to the proportions of 1996. The meaning of the category “C/C++” was not explained but apparently it means “C or C++”. I did not use statistical tests because the subsample sizes were not reported. Their results of the year 1996 are quite large when compared with the present research because their sample was for all kind of IT positions. Anyhow, their results show similar trends as the present research and match well in that respect.

Table 2. Proportions of C/C++, Cobol, and Visual Basic from 1989 to 1996.
Source: Athey and Plotnicki (1998, p. 75).

Year	**C/C++ (%)**	**Cobol (%)**	**Visual Basic (%)**
1989	16.7	27.9	0.0
1992	23.5	14.1	0.0
1993	31.0	17.0	6.1
1996	34.6	17.2	15.5

The report by Salary Services (2004a, pp. 224–229) presents the numbers of positions for each individual skill but no proportions. It was not possible to count the proportions because the subsample sizes were not reported, but I counted the ranks for 150 individual skills for the years

1999–2003 from the results that were presented in the report. In addition, I used the ranks for the ten most common skills from the third quarter of 2004 that were presented at the web page (Salary Services 2004b). However, the ranks are not presented here but only commented because the permission from Salary Services Ltd. is required before any part of the survey may be reproduced. I compared the ranks against the results of the present research. The results match well because according to the results of Salary Services, C++ has lost its position as the most required programming language to Java, and the rank of Cobol has decreased strongly.

Apparently, Java is historically similar with Cobol in that respect that only these two languages have achieved a very strong position (e.g., a proportion of 50%) in the job market. In case of Java, it took only five years to achieve this level. Cobol was released in 1960. According to the results by Todd et al. (1995), the proportion of Cobol was 41.7% in 1970 and 52.5% in 1975 for programmer positions. Thus, it took the maximum period of 15 years that Cobol achieved a very strong position. It is possible that Cobol achieved a very strong position even in a shorter period than 15 years because no research results from the 1960's were available during the present research.

4.4.3 Database skills

Next are presented the proportions of the most common database skills. It can be noticed that the proportions of Oracle and Microsoft SQL Server have increased strongly. In addition, the proportion of Sybase has increased but not as strongly as for Oracle and SQL Server. One might expect that the proportion of DB2 would have decreased as the proportion of platform category "Mainframe or midrange" has decreased (Section 4.4.1). However, this is not the case but the proportion of DB2 has been approximately the same during the period. Based on the Z test for proportions, the differences between the years 1990 and 2004 are statistically very significant ($p<0.001$) for Oracle, SQL Server, and Sybase.

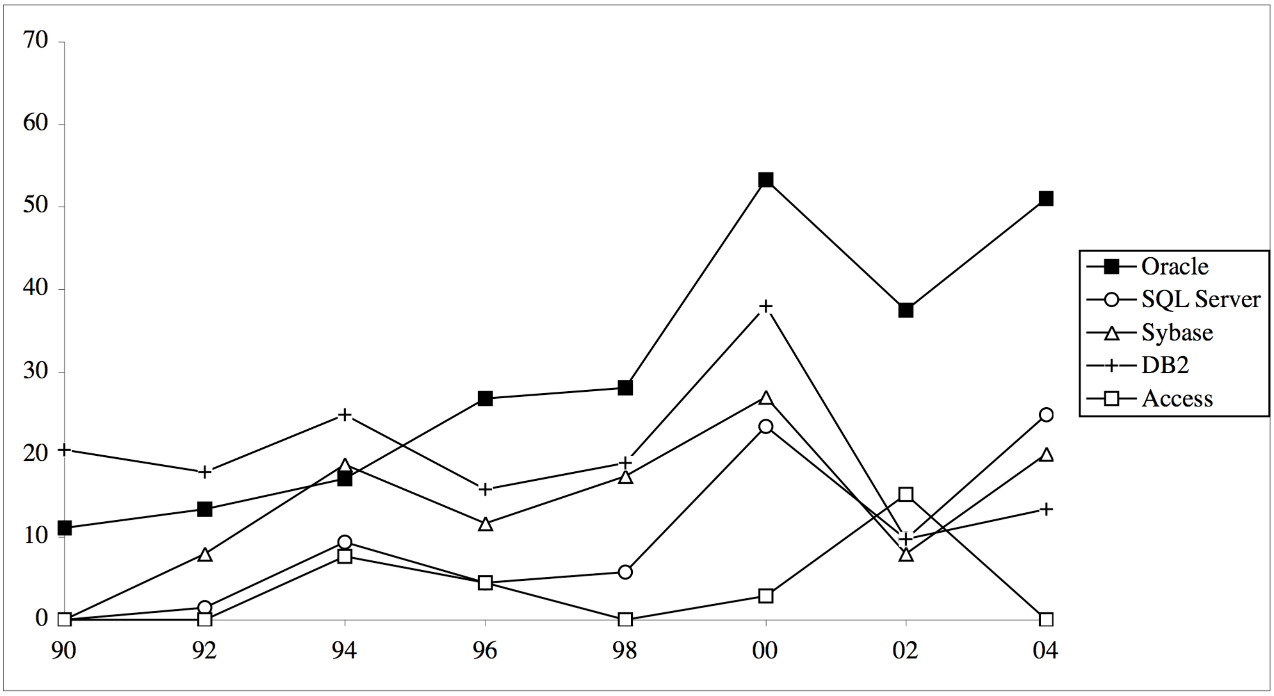

Figure 6. Proportions (%) of different database skills in years 1990–2004.

For brevity, the proportions of database languages such as SQL and PL/SQL are not presented as a figure. The most common database language skill was SQL and its proportions have varied from 4.3% to 20.5%. No trend could be noticed. The proportion was 7.4% in 1990 and 11.4% in 2004. Based on the Z test for proportions, this difference is statistically not significant ($p \geq 0.05$). The proportions of SQL for every year can be found in Appendix A.

Todd et al. (1995, p. 9) used only the category “Database” and Trower (1995, p. 599) the category “Relational DB” that are too general for comparison here. Maier et al. (1998, p. 41) used the following suitable categories: “DB2”, “Oracle”, and “Sybase.” Table 3 presents the sample sizes and proportions by Maier et al. for these three database skills. The columns are ordered according to the proportions of 1993/94. Based on the Z test for proportions, the differences of DB2 and Oracle between the years 1983/84 and 1993/94 as well as the difference of Sybase between the years 1988/89 and 1993/94 are statistically very significant ($p<0.001$). Their proportions of DB2 and Oracle in the years 1988/89 and 1993/1994 are smaller than in the present research during the period 1990–94 but this is reasonable because their sample included IT positions of all kind, not just software developers. Their proportion for Sybase in 1993/94 is a little smaller than in the present research in the year 1994. To sum up, their results show that the increase of DB2 and Oracle began in the early 1980’s, and the increase of Sybase in the early 1990’s. Generally, their results in 1993/94 match quite well with the results of the present research.

Table 3. Proportions (%) of DB2, Oracle, and Sybase from 1978/79 to 1993/94.
Source: Maier et al. (1998, pp. 38 and 41).

Years	Sample size	Sybase (%)	Oracle (%)	DB2 (%)
1978/79	1795	-	-	-
1983/84	1632	-	0.1	0.1
1988/89	1997	0.3	2.0	9.3
1993/94	1886	12.6	11.9	10.3

Dash (-) indicates that the result was not reported.

Athey and Plotnicki (1998, p. 78) used the following categories: Access, DB2, Informix, Oracle, Sybase, and “General SQL.” The results are presented in Table 4 where the columns are ordered according to the proportions of 1996. I did not use statistical tests because they did not report the subsample sizes. Their results match well with the results of the present research when it is considered that their sample was for all IT positions. The greatest differences are that the proportion of the category “General SQL” is greater than the proportion of the category “SQL” in the present research and the category “SQL Server” was not used or reported in their research.

Table 4. Proportions (%) of various database skills from 1989 to 1996.
Source: Athey and Plotnicki (1998, p. 78).

Year	Oracle (%)	General SQL (%)	Sybase (%)	DB2 (%)	Access (%)	Informix (%)
1989	4.3	3.2	0.5	7.9	0.0	1.9
1992	7.3	6.2	2.7	7.7	0.0	2.6
1993	17.4	3.6	12.4	10.0	1.9	5.2
1996	18.9	14.4	10.9	9.1	6.7	5.7

As for programming languages, I counted the ranks for database skills from the results by Salary Services (2004a; 2004b). This comparison was interesting, because the results by Salary Services included also the category "SQL Server" that was not used in the other longitudinal analyses. Also these ranks match quite well with the results of the present research because the order among database vendors is similar to the present research. The greatest difference was that according to Salary Services, SQL was the most common database skill and thus has apparently been required considerably more often than according to the present research.

As with Java, it is historically exceptional that Oracle has achieved such a strong position (i.e., the proportion of 50%). No other database product or vendor achieving this level was found during the literature survey of the present research. For example, based on previous job advertisement analyses, DB2 has never achieved the proportion of 50%. Apparently, the first commercially successful database management system was IBM's IMS. According to Maier et al. (1998), the proportion of IMS DMS/DB was at the greatest 16.5% in 1983/84. However, their result was not comparable with the results of the present research because their sample included all IT positions, not just software developers.

4.4.4 Networking skills

Next are presented the proportions of the most common networking skills in Table 5. The mean from proportions of the subsamples is presented on the last row. The columns are ordered according to the means. The category "network or networking" means no individual product or skill but the fact that the phrase "network" or "networking" was mentioned. No trends could be noticed. Statistical tests were not used to analyze the results because the differences between the years 1990 and 2004 are so small.

Table 5. Proportions (%) of different networking skills in years 1990–2004.

Year	TCP/IP (%)	Internet (%)	Novell (%)	network or networking (%)	LAN (%)
1990	1.6	0.0	1.1	0.0	2.1
1992	1.5	0.0	4.0	2.5	1.0
1994	3.4	0.0	1.7	2.6	3.4
1996	5.3	1.1	4.5	2.3	2.3
1998	2.5	5.8	0.8	0.0	0.0
2000	5.8	5.8	0.0	0.7	0.0
2002	8.0	4.5	1.8	2.7	0.9
2004	2.0	2.0	0.0	1.3	1.3
Mean	**3.8**	**2.4**	**1.7**	**1.5**	**1.4**

Maier et al. (1998, p. 41) is the only previous longitudinal research that contains suitable results for comparing trends. The results are presented in Table 6. They did not use "TCP/IP" or "Internet" as categories. The biggest differences are that the proportions of Novell and the phrase "communication" were considerable greater than in the present research. In the present research, the proportion of the category "communications" was 0.5–3.4% in 1990–96 and 0% in 1998–2004.

Table 6. Proportions (%) of networking skills from 1978/79 to 1993/94.
Source: Maier et al. (1998, pp. 38 and 41).

Years	Sample size	Communi-cation (%)	Network (%)	Novell (%)
1978/79	1795	3.0	0.1	-
1983/84	1632	3.7	0.9	0.4
1988/89	1997	7.3	0.7	3.6
1993/94	1886	5.6	0.6	12.8

Dash (-) indicates that the result was not reported.

Trower (1995, p. 597) reported that the most common networking skills sought in 1995 were Unix, Novell, TCP/IP, and WAN. In the present research, Unix was not classified as a networking skill but as an operating systems skill. His results about Novell and TCP/IP are similar to the present research.

Athey and Plotnicki (1998, p. 84) reported that the proportion of Novell was 13.3% in 1996 and 13% in 1993, and the proportions of Internet 14% and TCP/IP 7.6% in 1996. In addition, they used (p. 86) the categories "HTML" and "WWW" that were not used as networking skills in the present research. Their results for Internet and Novell are considerable greater than in the present research whereas the result of TCP/IP matches well enough.

In addition, I counted the ranks for networking skills from the results by Salary Services (2004a; 2004b). No general trend could be noticed because the ranks of some networking skills increased, some stayed at the same level, and some decreased. According to the ranks, the most common networking skill was clearly TCP/IP of which ranks stayed at the same level during the years 1999–2003. This matches well with the results of the present research. CISCO was usually ranked as the second common networking skill. This is different from the present research where CISCO was not among the top five networking skills and the proportions of CISCO were very low (0.0–0.7%). The ranks of other networking skills varied so much and were close to each other that it was hard to find the third common skill. Salary Services Ltd. apparently did not use Internet and phrases as "network" as categories.

To sum up, the need for individual networking skills has been relatively low or at most moderate during the period 1978–2004. In addition, it appears that Novell has lost its position as the most required networking skill to TCP/IP.

4.4.5 Distributed technology skills

For distributed technology skills, three categories were used: Microsoft, Sun, and Other. A position was classified in a certain category if at least one skill of the category was mentioned. It is possible that one position was classified to several categories. The skills of each category are presented in the following lists:

- Microsoft: .NET, Active X, ASP, DCOM, IIS, and MTS
- Sun: EJB, J2EE, JSP, RMI, and Servlets
- Other: technologies that do not belong in the previous two categories (e.g., applications server, client-server, CORBA, PowerBuilder, Tuxedo, Tibco, WebLogic, and WebSphere).

The proportions of these three categories are presented in Figure 7. As one can expect based on the results presented earlier in Section 4.3 about distributed technologies, the proportions for all three categories have increased strongly. In particular, it is interesting to compare the categories Microsoft and Sun because these are competing technology groups. It appears that Microsoft holds a stronger position than Sun because every year from 1996 to 2004 Microsoft's proportion was greater than Sun's proportion. Based on the Z test for proportions, the difference between Microsoft and Sun in the year 2004 is statistically almost significant ($p<0.05$). However, there is a statistically significant difference when the comparison is done over the period 1998–2004 and the subsamples from these years are combined. During this period, the overall proportion was 22.0% for Microsoft and 12.1% for Sun (n=519). Based on the Z test for proportions, this difference is statistically very significant ($p<0.001$). Thus, there was some evidence that skills related to Microsoft's products were required more often than skills related to Sun's products. However, in my opinion this evidence was not strong because according to (Surakka 2004, p. 5), Sun's proportion was a little greater than Microsoft's proportion (20% and 17%, respectively) but the difference there was not statistically significant. The sample (N=224) was from an American recruiting service Dice[2]. Thus, the results are contradictory.

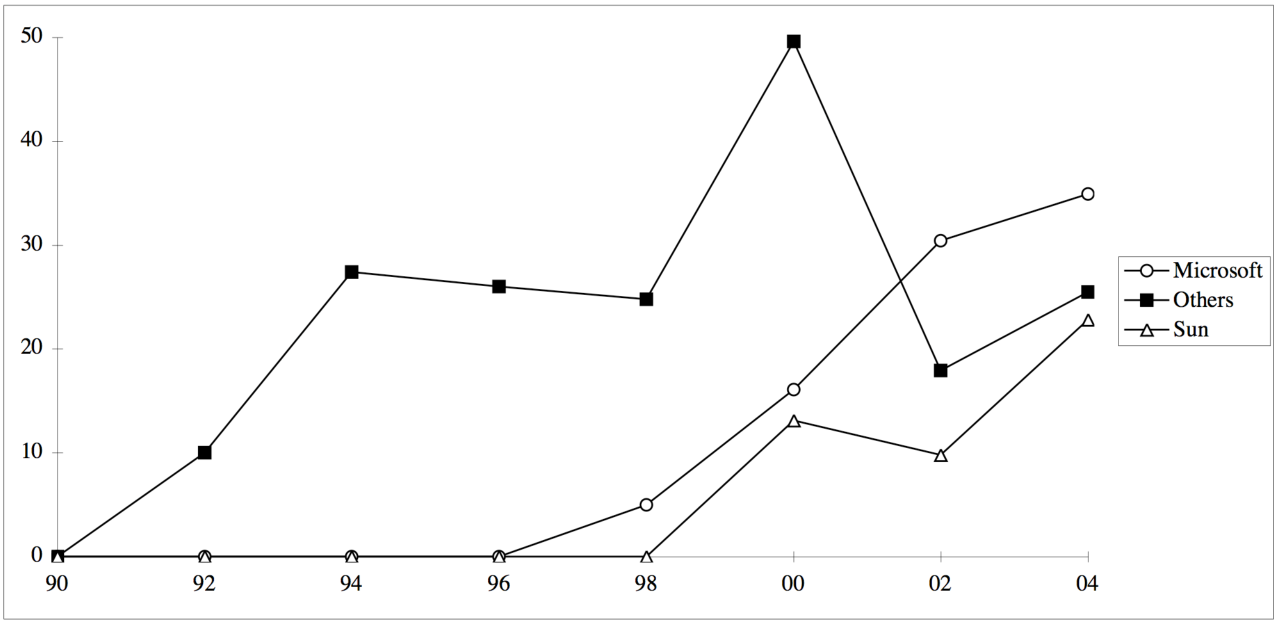

Figure 7. Proportions (%) of three categories related to distributed technology skills in years 1990–2004.

In addition, the proportions of individual distributed technology skills such as Java Server Pages and CORBA were counted. These results are shown in Appendix B.

The previous longitudinal analyses are not suitable for comparisons here. The category "Others" of the present research could be compared against the category "Client/Server" of Athey and Plotnicki (1998), and Trower (1995) but this has already been presented in Section 4.3. However, I counted the ranks for distributed technology skills from the results that were presented in the report (Salary Services 2004a). These results are compared against the present research in Appendix B.

2 In the USA, Dice (http://www.dice.com) is apparently the biggest web recruiting service in the field of IT. See Sections 4.5.1 and 5.1 for more information.

4.5 Miscellaneous results

In this section, some miscellaneous results are presented. These results were obtained for interpreting the results that are presented previously in Sections 4.1–4.4.

4.5.1 Different data source?

It is possible that for past five years or so, newspapers and professional magazines such as Computerworld are no longer a suitable choice as a single data source but a web recruitment service should be used instead (see Section 5.1). In this subsection, results from Computerworld and a web recruitment service Dice are compared against each other. In the USA, Dice is apparently the biggest web recruitment service in the field of IT. Both samples came from January 2004 and were coded manually. Only required skills were analyzed. The mean of the number of required skills was 8.89 for Computerworld (N=140, s=6.96) and 5.35 for Dice (N=223, s=3.51). Based on the Student t test and the Smith-Satterthwaite procedure, the difference between these two means is statistically very significant (p<0.001).

Next, it is calculated if the difference between the means from Computerworld in September 1990 and Dice in January 2004 is statistically significant. The means were 3.57 (n=189, s=3.63) and 5.35 (N=223, s=3.51), respectively. Based on the Student t test and the Smith-Satterthwaite procedure, the difference between these two means is statistically very significant (p<0.001). Thus, the number of required skills has increased during past 15 years also based on this comparison. However, the increase is a lot more moderate than based on data from Computerworld only.

In Table 7 it is presented the proportions for the same skills that have been reported previously in Sections 4.3 and 4.4. Inside each part of the table, the rows are ordered according to the column "Difference." As one can expect based on the means of the number of required skills, the proportions of Computerworld are usually greater than proportions of Dice. Approximately half of these differences are statistically significant. These results from Dice indicate that increase for various skills might be a lot more moderate if the data source would be changed from Computerworld to Dice. There are several explanations for this difference. Some of these are mentioned later in Section 5.1 (Evaluation of research).

Table 7. Proportions (%) for various skills in Computerworld and Dice, and difference of these two proportions. Both samples are from January 2004.

Skill	Computerworld (N=140)	Dice (N=223)	Difference
At least one:			
Database	74.3	37.2	37.1***
Distributed technology	45.0	30.5	14.5**
Programming language	64.3	54.3	10.0
Operating systems	46.4	41.7	4.7
Platform skills:			
DOS or Windows	52.9	37.2	15.7**
Unix	40.0	24.7	15.3**
Macintosh	0.0	0.0	0.0
Cross-platform	20.7	23.3	-2.6
Mainframe or midrange	16.4	20.6	-4.2
Programming language skills:			
Java	42.1	26.0	16.1**
Visual Basic	27.1	11.7	15.4***
Cobol	9.3	4.0	5.3*
C++	27.1	23.3	3.8
C	13.6	15.2	-1.6
Database skills:			
Oracle	55.0	22.0	33.0***
SQL Server	19.3	10.3	9.0*
Access	9.3	1.8	7.5***
DB2	9.3	4.9	4.4
Sybase	8.6	7.2	1.4
SQL	11.4	18.8	-7.4
Distributed technology skills:			
Sun	33.0	16.6	16.4***
Others	27.9	12.6	15.3***
Microsoft	17.9	14.8	3.1

One asterisk (*) refers to almost significant ($p<0.05$), ** to significant ($p<0.01$), and *** to very significant ($p<0.001$).

4.5.2 Different month?

Next, the results from Computerworld in January and September 2004 are compared against each other. The means were 8.89 (N=140, s=6.96) and 7.66 (n=145, s=5.67), respectively. Based on the Student t test and the Smith-Satterthwaite procedure, the difference between these two means is statistically not significant ($p \geq 0.05$).

Table 8 presents the proportions for these two samples. Inside each part of the table, the rows are ordered according to the column "Difference." It can be noticed that the proportions of September are typically greater than in January. Some of these differences are statistically significant. I have no good explanation for these differences. Interestingly, the proportions of January are for most items smaller even though its mean of the number of required skills was a little greater. Anyhow, also this comparison indicates that increase for various skills might be more moderate if some other sampling strategy would be used (e.g., every fifth year and four numbers per year).

Table 8. Proportions (%) for various skills in Computerworld in January and September 2004, and difference of these two proportions

Skill	**January (%) (N=140)**	**September (%) (N=149)**	**Difference (%)**
At least one:			
Distributed technology	45.0	59.7	14.7*
Programming language	64.3	70.5	6.2
Operating systems	46.4	51.7	5.3
Database	74.3	69.8	-4.5
Platform skills:			
DOS or Windows	52.9	64.6	11.7*
Mainframe or midrange	16.4	21.1	4.7
Unix	40.0	40.8	0.8
Macintosh	0.0	0.0	0.0
Cross-platform	20.7	9.5	-11.2**
Programming language skills:			
C	13.6	29.5	15.9**
Java	42.1	49.0	6.9
Visual Basic	27.1	32.2	5.1
C++	27.1	30.9	3.8
Cobol	9.3	3.4	-5.9*
Database skills:			
SQL Server	19.3	24.8	5.5
Sybase	8.6	20.1	11.5**
DB2	9.3	13.4	4.1
Oracle	55.0	51.0	-4.0
SQL	11.4	11.4	0.0
Access	9.3	0.0	-9.3***
Distributed technology skills:			
Microsoft	17.9	34.9	17.0**
Others	27.9	20.8	-7.1
Sun	33.0	22.8	-10.2

One asterisk (*) refers to almost significant ($p<0.05$), ** to significant ($p<0.01$), and *** to very significant ($p<0.001$).

4.5.3 Required experience

It is possible that the requirements have been increased, for example, because the level of required experience might have been increased in general. If more experience measured in working years would be required, it would be natural to require more individual skills as well.

Required years of experience were coded from those positions where the years were mentioned. Only samples from the years 1990 and 2004 were analyzed. If required years were not mentioned, the position was excluded from this analysis. If the position required a master's degree, this was calculated as two extra years. For example, the text "Masters in CS and 6 months experience" was coded as 2.5 years. In 1990, the mean of required years was 2.333 (s=0.801, n=42) and in 2004, the mean was 2.271 (s=1.292, n=131). Based on the Student t test and the Smith-Satterthwaite procedure, the difference between these two means is statistically not significant ($p\geq0.05$). Thus, this is not a likely explanation for the increased requirements.

5 Discussion

In this section, the research is evaluated, conclusions are drawn, implications to education are considered, and some possibilities for future research are presented.

5.1 Evaluation of research

The use of web recruiting services has increased strongly during the past five years and as a consequence, the number of newspaper advertisements has apparently decreased. For example, I estimated that Dice would publish approximately 0.3–0.8 million job advertisements and Computerworld approximately 3,000 job advertisements in 2004. In other words, currently the web recruiting services seem to dominate the IT job advertisement market in the USA. Thus, it is possible that Computerworld is not a representative source for the period 2000–2004. However, in my opinion, Computerworld should be a representative source for the period 1990–1998.

There are two other problems with the representativeness of the sample: (a) Only one number per year was chosen as the sampling strategy. As the results in Section 4.5.2 show, this was a questionable decision. (b) It is possible that some well-known companies such as HP, IBM, Microsoft, and Sun Microsystems do not advertise in Computerworld at all or only a little because interested job seekers search job advertisements directly from the web site of the company. In (Surakka 2004, pp. 8–9) it was found that Microsoft was announcing in Dice only a very small proportion of open positions when compared with the number of positions that were announced on Microsoft's own web site. Anyhow, a much bigger problem for the present research is the increased use of web recruiting services in general.

In Section 4.5.1, it was found that the proportions were often greater in Computerworld than in Dice. This difference is probably a consequence of several issues. Two possible contributory explanations are given next. I did not analyze either of these differences properly but just noticed them during the manual coding: (a) In Dice, it was apparently more common that skills were divided into required skills and desirable skills. This might decrease the results of Dice when compared with Computerworld because only the required skills were coded. A possible explanation for this difference is that Dice's instructions for advertisers might be different than in Computerworld. (b) It appears that in Dice the proportion of contractor positions was greater than in Computerworld. In Dice, it was compulsory to announce the type of a position (e.g., permanent or contractor). In Computerworld, this information was not usually given. If a contractor was typically searched for only a single project lasting, for example, 3–12 months, the number of required skills might be more limited than for a permanent full-time position.

Obviously, also the research method has its limitations. In some cases, job advertisements do not contain enough suitable information. Simply put, one can find out from job advertisements that some skills are probably important but solving out if some particular skill or subject is not important can be much more difficult. Analysis appears to work quite well for language and product names such as Java and WebSphere. For search phrases of other kind–that are typically more general terms such as "theory of computation" or "data structures and algorithms", situation is much more difficult. Other research methods can be used to gather information about the importance of these subjects and skills. For example, Lethbridge (2000, p. 46) found in his survey that the respondents considered "Data Structures" and "Software Architecture" as being important. Third example might be concurrent programming because, for example, Java can be used for both object-oriented and concurrent programming but it is usually not possible to find out from the job advertisements if concurrent programming is necessary. According to my unpublished Delphi study, a small group of experienced software developers (*N*=11) considered concurrent programming as being important: the mean was 3.1 on the scale 1 Not at all important … 4 Very important.

5.2 Conclusion and implications to education

Based on the results of the present research, the results by Gallivan et al. (2002), Maier et al. (1998), and Todd et al. (1995), the conclusion is that the technical requirements for software development positions have changed during the past 35 years so that the number of required individual skills has increased on average. This change is so large and found in more than one research that problems with sampling or other simple explanations are not plausible.

Duties of software developers have changed as technically more versatile in that respect that typically it is no longer enough to have skills only in 1–2 programming languages. In particular, the results in Section 4.3 indicate that it has become more and more common that software developers are expected to have database and distributed technology skills as well. For universities and other institutes, one possible strategy or tactic to respond to the increased and more complicated technical demands of industry is specialization to the certain skill groups or sectors of the job market.

A typical or reasonable list of skills for a graduate from a Bachelor program in computer science might be like the following when individual technical skills similar to the present research would be used: C, Java, SQL, UML, and Unix; that is, five skills altogether. For example, it is not reasonable to expect that graduates from an undergraduate program would have good skills in more than two programming languages. For a master degree, a similar list could be, for example, the following: C, Java, JDBC, RMI, Scheme, SQL, UML, and Unix; that is, eight skills. 5–8 skills are still at the same level as the number of required skills in the job advertisements on average in the year 2004. However, if the number of required skills continues to increase, degree programs might have severe difficulties in following this change. After all, degree requirements typically include many other topics than technical skills as well (e.g., mathematics, general skills such as project management, communication and language skills).

McCauley and Manaris (2002) reported that in ABET/CAC accredited bachelor programs in computer science the five most common programming languages that were taught first during the academic year 2001–2002 were Java (49%), C++ (40%), C (11%), Ada (4%), and Pascal (2%). In this respect, the match between curricula and the job market was good or at least satisfactory because the order in the job advertisements in the year 2004 was Java, Visual Basic, C, C++, and Cobol (Section 4.4.2). This comparison does not imply that all degree programs should use, for example, only the three most common languages. There can be other reasons than the popularity of language in industry to choose the programming language used in education. For example, some IS/IT programs might specialize in the certain part of the job markets (e.g., mainframe environments) and use Cobol. I found no similar statistics about the use of programming languages from IS/IT degree programs. However, according to the Accreditation Board for Engineering and Technology (2004), the number of accredited Information Systems programs is only 11 whereas the number of accredited Computer Science programs is 206. Thus, only 5% of accredited CS/IS programs are IS programs. It is also likely that the number of students in IS/IT programs is a lot smaller than in CS programs. Thus, the overall match appears as being good if it is assumed that IS/IT programs use Cobol more often than CS programs.

In addition, McCauley and Manaris (2002) reported which upper-level courses were offered, required, and offered as elective during the academic year 2001–2002. The course Operating

Systems was required in 96%, Database Management Systems in 31%, and Networks in 18% of the departments. This matches well with the results of the present research in relation to that operating systems and database skills were required in job advertisements more often than networking skills. However, the proportion of the course Database Management Systems could be considered low when compared how often database skills were required in job advertisements. I repeat the suggestion that I made already in (Surakka 2004, p. 8): The course Database Management Systems should be made compulsory more often. I do not suggest a course on distributed technologies as compulsory in undergraduate programs because based on my earlier findings (ibid., p. 6) distributed technology skills are required in entry-level positions less often than in senior-level positions. The proportion of at least one distributed technology skill was 27% for entry-level positions (n=41). A course on distributed technologies is very suitable as compulsory or at least as elective for master's programs in certain specializations (e.g., Software Systems).

5.3 Future research

At least in the USA and UK, the web recruiting services are now dominating IT job advertisement markets. This is a problem for longitudinal research setting because the data of the web services of the past few years are not as easily available as major newspapers and professional magazines are. Already now, it is difficult to conduct a reasonable longitudinal research if this problem is not solved. The problem is so essential for longitudinal job advertisement analyses that most of this subsection considers only the use of data from web recruiting services. In addition, one actual suggestion for future research is mentioned later in Section 5.3.2.

5.3.1 Use of data from web services

Previously, Litecky and Arnett (2001) have written about the need to have a job skills database available for researchers. In 2001, they submitted a proposal for the National Science Foundation about a job skills database that would be suitable for longitudinal research. Unfortunately, they did not get funding and therefore did not develop a database.

Organizations such as Information Technology Association of America (ITAA) or the ACM Special Interest Group on Management Information Systems (SIGMIS) might have better opportunities than individual researchers or research groups to make a long-range agreement about the deployment of databases of web recruiting services. From the viewpoint of a web recruiting service, it might be a problem if 5–20 individual researchers or research groups ask different kind of copies of the database every year.

Technically, it would be possible to make a copy of a web recruiting database using, for example, a script program that would automatically copy all advertisements. However, without permission this is probably illegal. At least Dice (2004, Section "Code of conduct") forbids the use of such programs without prior written authorization: "Use any robot, … or automatic device or process to … reproduce … its contents."

5.3.2 Concentrating on internship and entry-level positions?

Educators have conducted most job advertisement analyses. In the future, it would be interesting to conduct a research that concentrates on internship and/or entry-level positions because these are the most relevant to graduates from a Bachelor program. During the literature search of the present research, no such analysis was found. In (Surakka 2004, p. 6) one section considered differences between entry-level and senior-level positions but the whole research was not concentrated on entry-level positions.

Acknowledgements

I would like to thank professor L. Malmi from the Helsinki University of Technology and professor emeritus V. Meisalo from the University of Helsinki for commenting on the manuscripts of the present report.

References

Accreditation Board for Engineering and Technology, 2004. Retrieved December 9, 2004, from the Accreditation Board for Engineering and Technology web site: http://www.abet.org/.

Athey, S. and Plotnicki, J., 1998. The evaluation of job opportunities for IT professionals. Journal of Computer Information Systems (Spring), 71–88.

Dice, 2004. Terms of Use Agreement. Retrieved on November 2, 2004, from Dice's web site: http://employer.dice.com/employer_r.epl?op=32.

Gallivan, M., Truex III, D.P., & Kvasny, L., 2002. An analysis of the changing demand patterns for information technology professionals. SIGCPR'02, May 14–16, 2002, Kristiansand, Norway, 1–13.

Institute of Electrical and Electronics Engineers, 1990. IEEE standard glossary of software engineering terminology. IEEE Std 610.12–1990.

Lethbridge, T., 2000. What knowledge is important to a software professional? Computer (May), 44–50.

Litecky, C. & Arnett, K.P., 2001. An update on measurement of IT job skills for managers and professionals. Proceedings of the Seventh Americas Conference on Information Systems, Boston, MA, August 2001, 1922–1922.

Litecky, C., Prabhakar, B., & Arnett, K., 1996. MIS job market: Shaken but not stirred. Journal of Systems Management, 47, 4 (July/August), 51–54.

Maier, J.L., Clark, W.J, & Remington, Jr., W.S., 1998. A longitudinal study of the management information systems (MIS) job market. Journal of Computer Information Systems (Fall), 37–42.

Milton, J. S. & Arnold, J. C. (2003). Introduction to probability and statistics. 4th ed. McGraw-Hill, New York.

Nakayama, M. & Sutcliffe, N., 2000. Introduction to research on IT skill issues. Proceedings of Americas Conference on Information Systems 2000, August 10–13, Long Beach, CA. Retrieved May 19, 2004, from The Association for Information System Electronic Library Web site: http://aisel.isworld.org/password.asp?Vpath=AMCIS/2000&PDFpath=376.pdf, 1930–1934.

Nakayama, M. & Sutcliffe, N.G., 2001. IT skills portfolio research in SIGCPR proceedings: Analysis, synthesis and proposals. Proceedings of the 2001 ACM SIGCPR conference on Computer personnel research, San Diego, CA, 100–113.

Prabhakar, B., Litecky, C., & Arnett, K., 1995. Boom times ahead! Journal of Systems Management (1), 24–28.

Salary Services Ltd., 2004a. ComputerWeekly. Survey of appointments data & trends. Quarterly survey (January).

Salary Services Ltd., 2004b. TOP IT Skills–All. Retrieved December 27, 2004, from the Salary Services Ltd. web site: http://www.salaryservices.co.uk.

Surakka, S., 2004. Analysis of job advertisements: What technical skills do software developers need? In: Korhonen, A. & Malmi L. (eds.). Kolin Kolistelut—Koli Calling 2004. Proceedings of the Fourth Finnish/Baltic Sea Conference on Computer Science Education, October 1–3, 2004, Koli, Finland, pp. 47–56.

Todd, P.A., McKeen, J.D., & Gallupe, R.B., 1995. The evolution of IS job skills: A content analysis of IS job. MIS Quarterly (March), 1–27.

Trower, J.K., 1995. The impact of job skill requirements on I.S. curricula. Proceedings of the first Americas conference on Information Systems. August 25–27, 1995, Pittsburg, PA, 597–599.

Ulrich's periodical directory, 2004. Retrieved November 24, 2004, from the web site: http://www.ulrichsweb.com/ulrichsweb.

Appendices

Appendix A: Results as tables

In this appendix, the results of Figures 1–7 are presented as tables. Table A.1 corresponds Figure 1 and so on. However, some tables contain more results that the corresponding figures.

Table A.1. Proportions (%) of four job titles in years 1990–2004.

Year	*n*	Programmer (%)	Software developer (%)	Software engineer (%)	Systems analyst (%)
1990	189	85.2	2.1	9.5	3.2
1992	201	52.7	12.4	32.3	2.5
1994	117	58.1	31.6	3.4	6.8
1996	265	51.7	27.9	17.0	3.4
1998	121	26.4	38.8	27.3	7.4
2000	137	24.8	35.0	32.8	7.3
2002	112	47.3	10.7	38.4	3.6
2004	149	36.9	14.8	43.0	5.4

Table A.2. Means (*M*) of number of required technical skills for all software developer positions, programmers, and others than programmers in years 1990–2004. In addition, sample sizes (*n*) and standard deviations (*s*) are presented.

Year	All developer positions			Programmers			Others than programmers		
	n	*M*	*s*	*n*	*M*	*s*	*n*	*M*	*s*
1990	189	3.57	3.63	161	3.37	3.66	28	4.75	3.31
1992	201	3.42	3.72	106	4.11	3.79	95	2.65	3.50
1994	117	5.73	7.41	68	5.32	7.07	49	6.29	7.90
1996	265	4.83	5.09	137	4.28	4.52	128	5.41	5.59
1998	121	5.10	5.13	32	4.63	5.05	89	5.27	5.18
2000	137	7.22	8.03	34	5.59	8.56	103	7.76	7.82
2002	112	7.05	4.20	53	7.40	3.86	59	6.75	4.49
2004	145	7.66	5.67	53	8.77	5.77	92	7.01	5.55

Table A.3. Proportions (%) of at least one skill in four skill categories in years 1990–2004.

Year	*n*	Programming language (%)	Database (%)	Distributed technology (%)	Operating system (%)
1990	189	48.7	25.9	0.0	28.0
1992	201	43.8	36.8	3.5	41.8
1994	117	58.1	39.3	17.9	48.7
1996	265	59.6	51.7	14.7	43.4
1998	121	57.9	55.4	22.3	42.1
2000	137	75.9	62.8	54.0	40.9
2002	112	80.4	70.5	50.9	41.1
2004	149	70.5	69.8	59.7	51.7

Table A.4. Proportions (%) of different platforms in years 1990–2004.

Year	*n*	DOS or Windows (%)	Unix (%)	Mainframe or midrange (%)	Cross-platform (%)	Macintosh (%)
1990	189	4.8	7.4	74.1	7.9	0.0
1992	201	13.4	17.4	54.7	10.0	1.0
1994	117	41.0	30.8	56.4	8.5	3.4
1996	265	32.5	26.8	39.2	21.1	2.3
1998	121	35.5	21.5	40.5	24.0	0.0
2000	137	62.0	33.6	38.0	18.2	0.0
2002	112	67.0	26.8	13.4	11.6	0.0
2004	149	64.6	40.8	21.1	9.5	0.0

Table A.5. Proportions (%) of different programming language skills in years 1990–2004.

Year	*n*	Java (%)	Visual Basic (%)	C++ (%)	C (%)	Java-Script (%)	Perl (%)	Visual C++ (%)	C# (%)	VBScript (%)	Cobol (%)
1990	189	0.0	0.0	4.8	15.3	0.0	0.0	0.0	0.0	0.0	40.7
1992	201	0.0	1.5	7.5	21.9	0.0	0.5	0.0	0.0	0.0	24.4
1994	117	0.0	18.8	29.1	24.8	0.0	0.0	3.4	0.0	0.0	29.1
1996	265	6.8	14.3	29.4	32.1	1.5	1.5	4.2	0.0	0.0	26.8
1998	121	9.9	13.2	18.2	27.3	0.0	1.7	10.7	0.0	0.0	23.1
2000	137	48.2	32.1	43.1	31.4	6.6	8.8	13.9	0.0	5.8	35.0
2002	112	51.8	50.0	28.6	17.0	10.7	1.8	4.5	2.7	5.4	18.8
2004	149	49.0	32.2	30.9	29.5	12.8	8.1	8.1	7.4	5.4	3.4

Table A.6. Proportions (%) of different database skills in years 1990–2004.

Year	*n*	Databases					Database languages	
		Oracle (%)	SQL Server (%)	Sybase (%)	DB2 (%)	Access (%)	SQL (%)	PL/SQL (%)
1990	189	11.1	0.0	0.0	20.6	0.0	7.4	0.0
1992	201	13.4	1.5	8.0	17.9	0.0	9.0	0.0
1994	117	17.1	9.4	18.8	24.8	7.7	4.3	0.0
1996	265	26.8	4.5	11.7	15.8	4.5	8.3	3.0
1998	121	28.1	5.8	17.4	19.0	0.0	18.2	0.8
2000	137	53.3	23.4	27.0	38.0	2.9	14.6	5.1
2002	112	37.5	9.8	8.0	9.8	15.2	20.5	8.0
2004	149	51.0	24.8	20.1	13.4	0.0	11.4	7.4

Table A.7. Proportions (%) of three categories related to distributed technology skills in years 1990–2004.

Year	*n*	Microsoft (%)	Sun (%)	Others (%)
1990	189	0.0	0.0	0.0
1992	201	0.0	0.0	3.5
1994	117	0.0	0.0	17.9
1996	265	0.0	0.0	14.7
1998	121	5.0	0.0	19.0
2000	137	16.1	13.1	41.6
2002	112	30.4	9.8	16.1
2004	149	34.9	22.8	20.8

Appendix B: Results from individual distributed technology skills

In Table B.1, the proportions of individual distributed technology skills in the years 1990–2004 are presented. The ten most common skills in 2004 were selected in the table. The columns are ordered according to the proportions in 2004. In addition to these ten skills, the phrase “client/server” was mentioned in the job advertisements starting from the year 1992. However, its proportion was not large enough in 2004 to be included in the table. The proportion of “client/server” was 5.0% in 1992 and after that it has varied from 4.7% to 17.9%. The maximum value 17.9% was from the year 1994 and the minimum value 4.7% from the year 2004.

Table B.1. Proportions (%) of individual distributed technology skills in years 1990–2004.

Year	*n*	ASP (%)	J2EE (%)	EJB (%)	JSP (%)	.NET (%)	Servlets (%)	DCOM (%)	CORBA (%)	WebLogic (%)	Power-Builder (%)
1990	189	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1992	201	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.5
1994	117	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.8
1996	265	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	15.8
1998	121	0.0	0.0	0.0	0.0	0.0	0.0	1.7	6.6	0.0	6.6
2000	137	12.4	1.5	5.1	7.3	0.0	9.5	0.0	12.4	2.9	18.2
2002	112	27.7	7.1	0.9	3.6	4.5	2.7	4.5	6.3	5.4	6.3
2004	149	28.2	14.1	11.4	11.4	10.7	10.7	7.4	6.7	6.7	6.0

The previous longitudinal researches do not contain results that would be suitable for comparison. However, the commercial report and the web page by Salary Services (2004a; 2004b) contain suitable results for comparison. I counted the ranks for distributed technology skills as previously for programming language and database skills. These ranks confirm that the importance of distributed skills has increased during the past five years. Also according to these ranks, Microsoft’s technologies are required more often than Sun’s (see Section 4.4.5). However, this comparison was problematic because Salary Services apparently did not use the category “J2EE” that was the most often mentioned Sun’s distributed technology skill in the present research.