Employment Analysis of ERD vs. UML for Data Modeling

Hsui-lin Winkler
Seidenberg School of Computer Science and Information Systems
Pace University
1 Pace Plaza, NYC, NY 10038
hwinkler@pace.edu

Henning Seip
SkillPROOF Inc.
nValley Technology Center, 2nd Fl.
470 Nepperhan Ave., Yonkers, NY 10701-6601
info@skillproof.com

ABSTRACT

In this paper we combined keyword search and content analysis to investigate the employment demands of ERD vs. UML in the context of data modeling. We found that data modeling is one of the popular skills required by many of the general business information systems professionals without specific methodology attached to the employment requirements. ERD and UML are required in narrower and often separate fields of building data models for information systems, in which ERD is mentioned in the domain of system analysis and design and database administration, while UML is required mostly in software engineering and application development. The division between ERD and UML in professional data modeling remains, although there is a small number of employment requirements that demand both methodologies and require a mapping capability between the two approaches.

Keywords: IT employment data, keyword search, content analysis, ERD, UML, data modeling

1. INTRODUCTION

Information and communication technology is one of the fastest changing fields in employment skills. This has resulted in constant revising of the academic curricula and textbooks to best match the education objectives and demands in professional skills (Gorgone, et al. 2002). In IT education, the various correlations of the recent IT employment opportunities and college students entering this field further stress the needs for examining the employment skills. The available employment opportunities published on the Internet have provided not only a job search tool for many job seekers but also a systematic monitoring of changes in employment skill demands in real-time (SkillPROOF Inc. 2004).

In this study, we analyzed IT employment data published by employers and focused on the ERD(Entity Relationship Diagram) vs. UML(Unified Modeling Language) requirements in the field of data modeling. We hope that our findings can supplement education decisions on what we should include in our teaching scopes and identify trends in the required professional skills

2. EMPLOYMENT DEMANDS

The employment data used in this study is published by individual companies on the Internet and collected daily by SkillPROOF Inc. since the beginning of 2004. The data is collected from up to 137 IT-focused companies. Each data sample contains attributes of company industry, posting date, job title, job responsibility, skill and education or training requirements. The general and background information of the data collection and categorization can be found on the website of SkillProof.com (SkillPROOF Inc. 2004).

From the archived data, there are total of 35,932 jobs. The job counts from the top 12 industries (among a total of 46 industries) are plotted in Figure 1 to illustrate the overall distribution of the employment demands. The distribution reflects the post dot.com and post Sept/11 IT employment demands.

3. DATA ANALYSIS

We first applied keyword search to the job description to categorize the relevant ERD vs. UML skill requirements
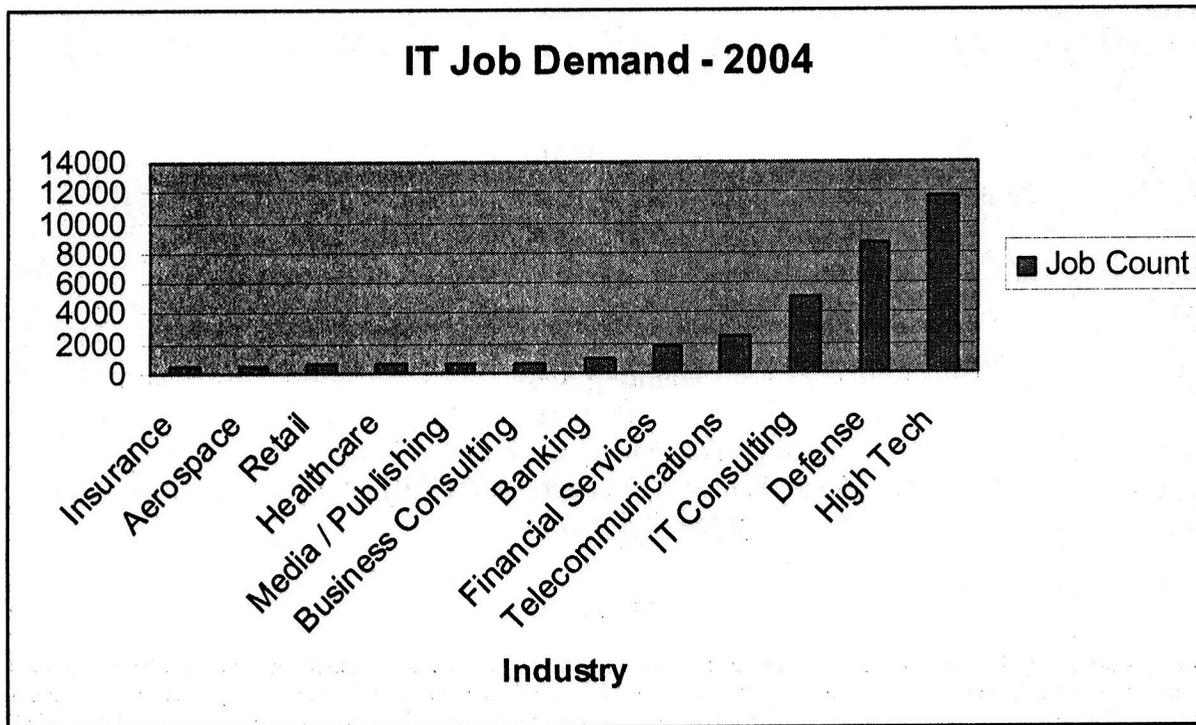
Figure 1. IT Job Counts Distribution for the Top 12 Industries.

according to industry and then according to job types or job functions. We further sampled the contents of job description to investigate the implications for the job requirements in ERD vs. UML.

3.1 Keyword Categorization According to 'Industry'

We used the keywords of 'data model' or 'data modeling', 'ERD' and a few commonly referenced design tools like ErWin(2006), Visio(2006) and a database tool 'TOAD'(2006) to search for data modeling and database analysis, design and management related employment requirements. Similarly, we used the keyword 'UML' to search through the same data sets. We classified the search results according to the top 12 industries identified in Figure 1 and plotted the job counts in Figure 2(a), (b) and (c) below for comparison. One extra industry 'pharmaceutical' is added because the job counts in that industry are within the range of interest in the new aggregation.

Both ERD and UML keyword searched job counts are notably reduced to about half compared with data modeling searched job counts. Overall, the distribution for data modeling and ERD are similar in that both job spectra are broadly distributed across the industry line. They both share three common demanding industries: defense, high-tech and IT consulting. One exception is that the retailing industry appears to be very significant in requiring data modeling skills whereas the telecommunication is an industry requiring significant employment of ERD skills.

For the skill demand in UML, the outcome is significantly different. The defense industry is very noticeable in its relatively large job number in requiring UML skill. In terms of the job counts, the demands for UML are similar to the demands in ERD in both high-tech and IT consulting. However, they are both less than 25% of those counted in the defense industry.

3.2 Keyword Categorization According to 'Job Type'

Another classification use keyword search is to sort all the job requirements according to the job function or job title. Use a standard job classification, we were able to use keywords of 'data modeling', 'ERD' or 'UML' to plot the searched skills vs. the types of jobs defined. The search results are summarized in Figure 3(a), (b) and (c) respectively for 'data modeling', 'ERD' and 'UML'. In this classification, we find that, again, UML has a lone popular job type as 'software development' whereas 'data modeling' or 'ERD' skill demands are more evenly distributed among the various job types, such as business analyst, software developer, IT consultant, technical writer and database administrator. One exception is that the job type of 'project manager' shows a significant number in ERD searched results but is not as popular in the 'data modeling' group.

Since the job type or job function is closely tied to the job skills, we did further analysis to investigate the combined skill requirements. For example, in data modeling keyword search we would add 'ERD' and/or 'UML' to 'data model-
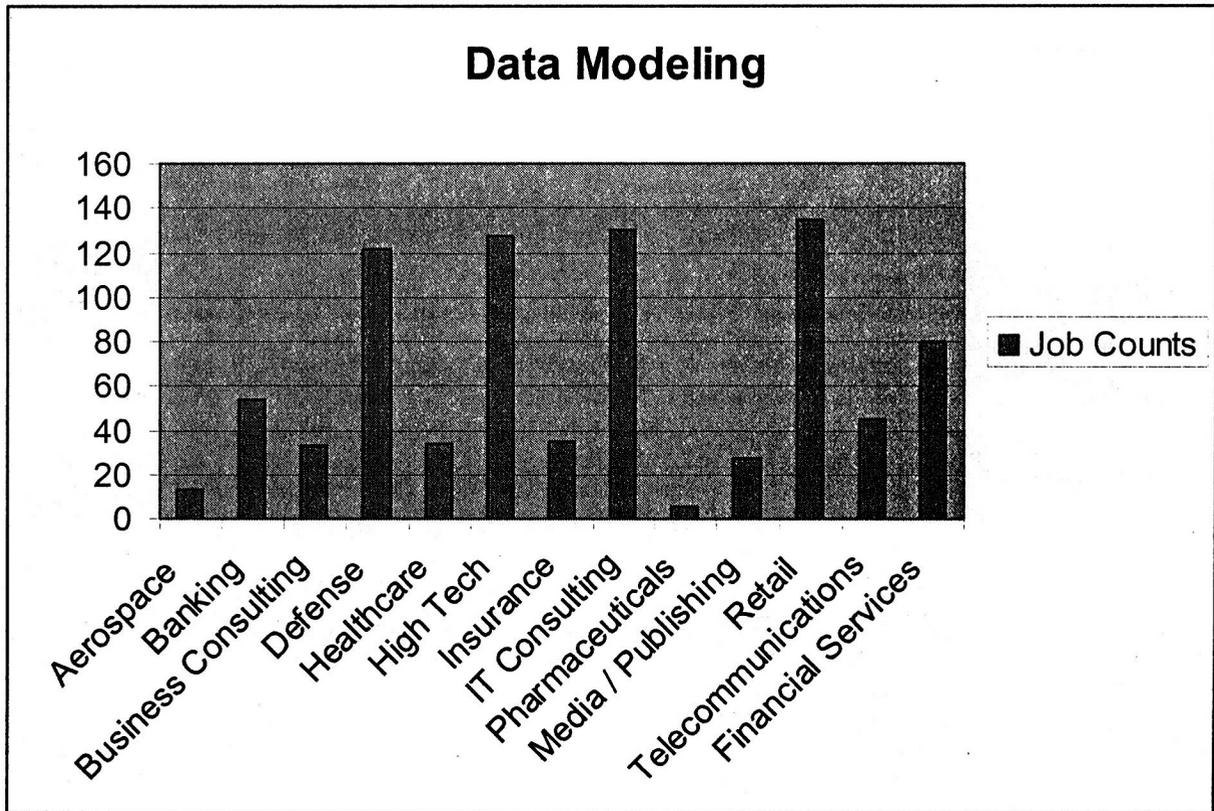
66

Data Modeling

Figure 2(a). Job Counts Containing Keywords 'Data Model' or 'Data Modeling' for the Top 13 Industries.
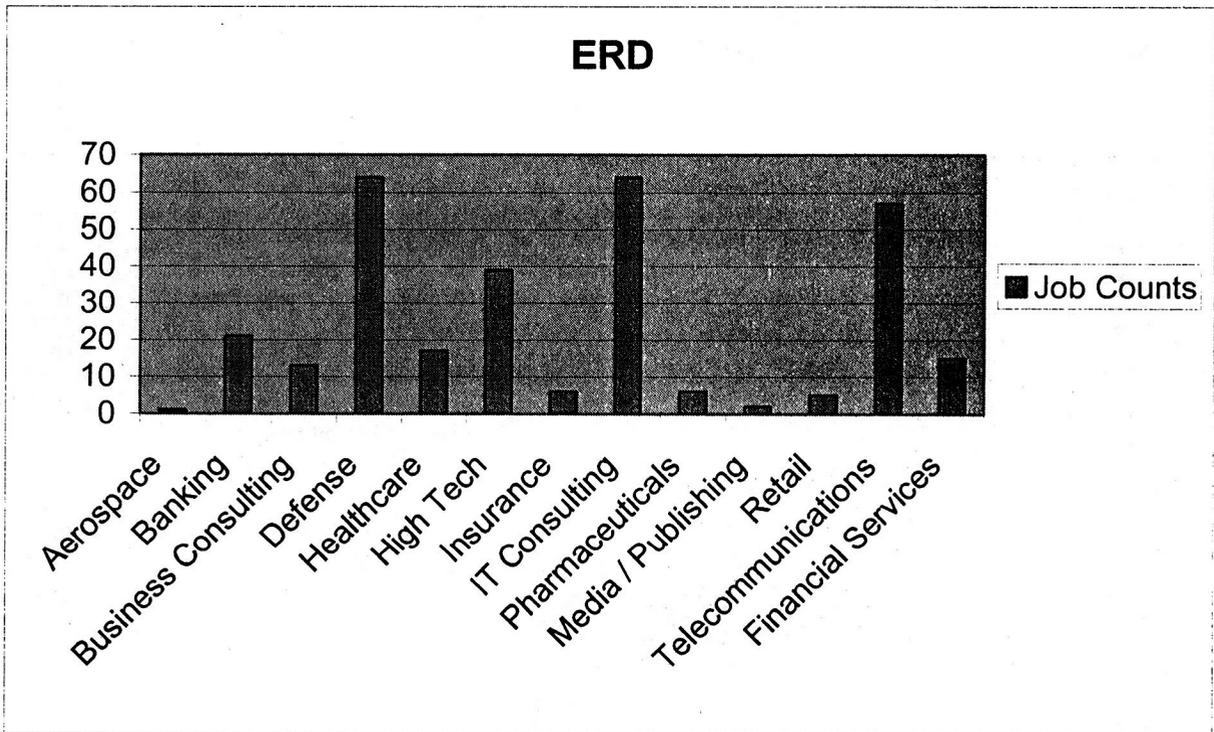
ERD

Figure 2(b). Job Counts Containing Keywords 'ERD', 'ErWin', 'Visio' or 'TOAD' for the Top 13 Industries.
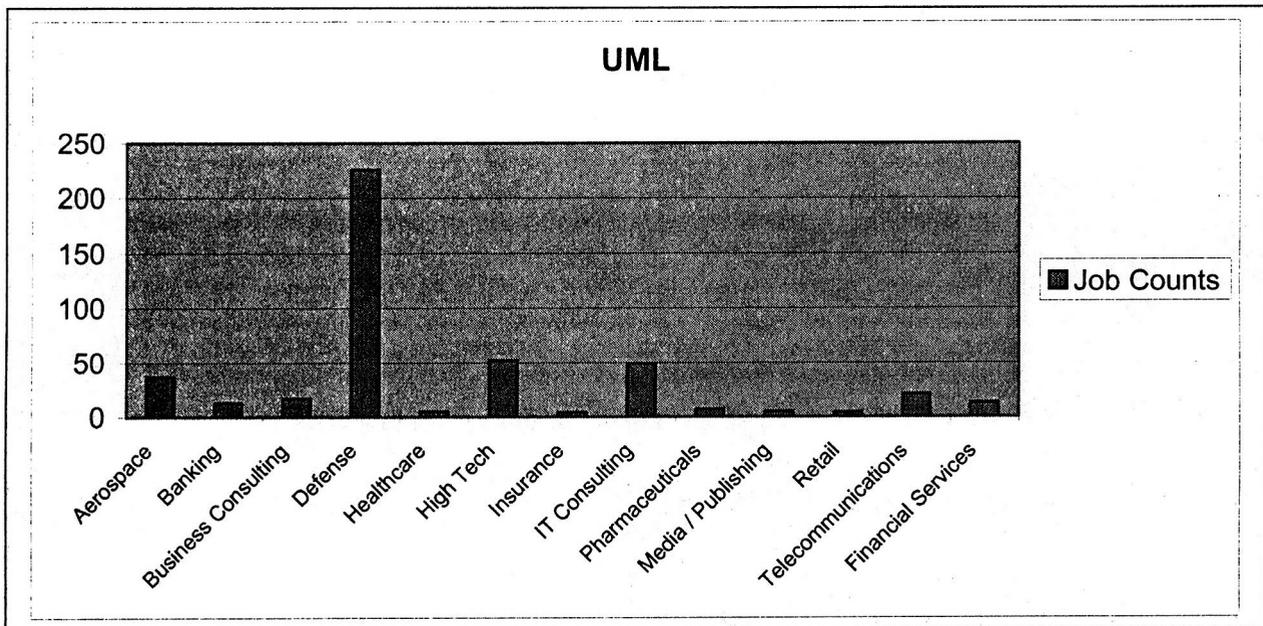
67

Figure 2(c). Job Counts Containing Keyword 'UML' for the Top 13 Industries.

ing'. Table 1 summarizes the resulting job numbers in both single and combined keyword search results. The numbers are somewhat surprising at first glance. It implies that a lot of job descriptions would mention a required or desired skill of data modeling without specifying any methodology. For example, a single keyword 'data modeling' search resulted in job counts of 867. Likewise, there are many ERD(318) or UML(457) search results without the reference of data modeling. When we use combined keywords of 'data modeling' and 'ERD', the resulting job counts are reduced to 56.

We therefore turned to content analysis to provide further insights and answers into the keyword search results.

Employment Skills Use the Following Keyword(s)	Number of Job Advertised
Data Modeling	867
ERD(and ERD Tools)	318
UML	457
Data Modeling and UML	39
Data Modeling and ERD	56
ERD and UML	9
Data Modeling and ERD and UML	2

Table 1. Job Counts from Keyword(s) Searched Results.

3.3 Sampled Content Analysis
In order to explain the keyword search results, we further examine the contents of the job description to analyze the details of the requirements. Due to the large volume of data involved, we select only one-week data here for presentation (the first week of December 2004).

In Table 2 we show one employment description each for the group selected as described in Table 1. In compiling the following comparison, we include only the job title, required skills and desired skills taken directly from the job description posted by individual companies.
From the above skill requirements description, we found that 'data modeling' skill is embedded more often in general business information systems skills than in the specific data analysis or modeling skills we were searching for. Although we found a very high number of job counts in data modeling, many do not require the specific methodologies of either ERD, ERD specific tools or the standard modeling technology like UML.

It is interesting to note that although the number distribution shown in Table 1 may have been unexpected, our findings are similar to what are reflected in the college textbooks used in the general discipline of information systems (i.e. O'Brien and Marakas, 2005; Whitten and Bentley, 1998). In many business information systems, the main contents are focused on enterprise information systems and management strategy, data analysis and modeling is included as one of the topics. The searched result using keyword 'data modeling' reflects basically this type of employment skill requirements.

When we use keyword like 'ERD' or related tools of data modeling, we begin to find the traditional discipline of system analysis and design, in which process modeling and data modeling are key elements in the contents (i.e. Hoffer, 2005; Kroenke, 2005; and Shelly, 2003). Both ERD and UML are usually included as the data modeling methodology, but if we focus on the data modeling and relational model, ERD is still the main approach and UML is often included in the later chapter or in an appendix.
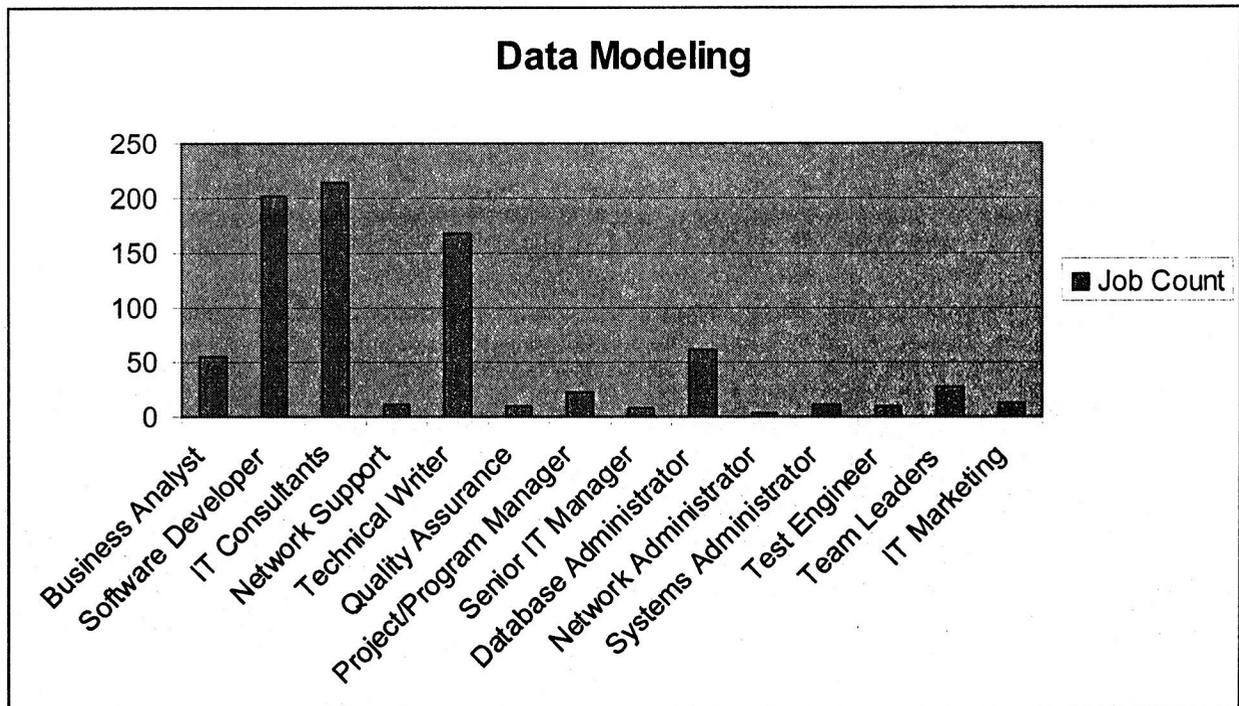
Data Modeling

Figure 3(a). Job Counts Containing Keyword 'Data Modeling' for the Top 14 Job Types.
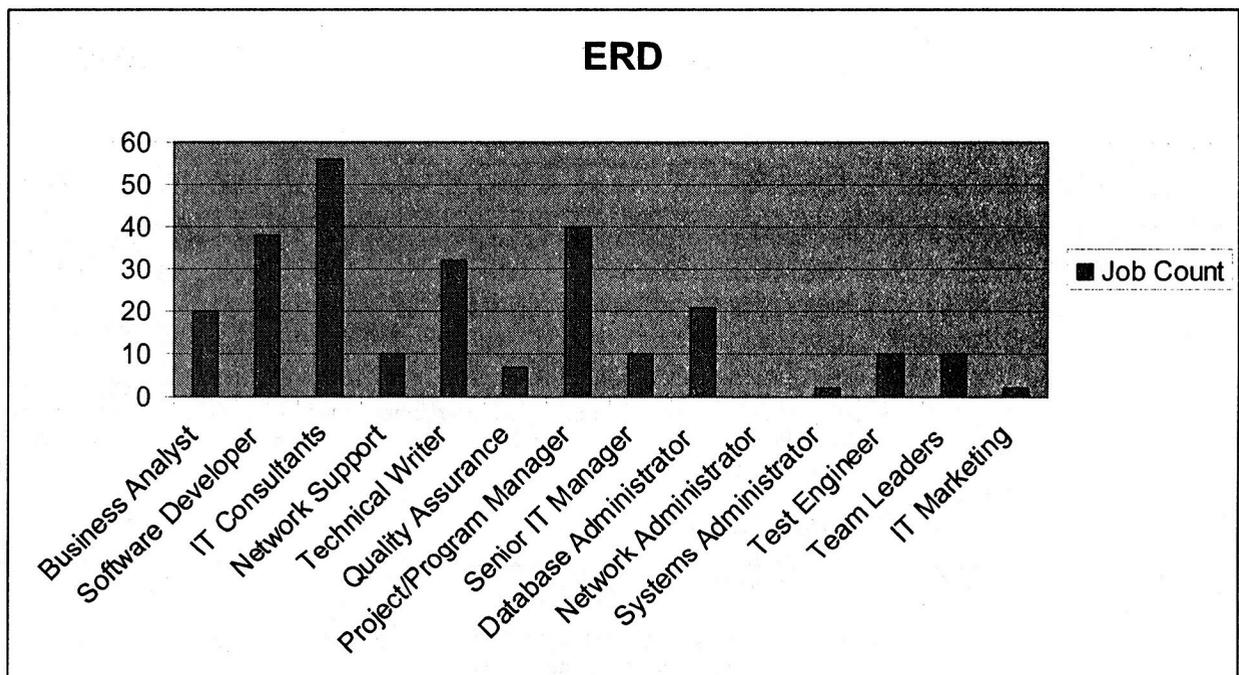
ERD

Figure 3(b). Job Counts Containing Keyword 'ERD', 'ErWin', 'Visio' or 'TOAD' for the Top 14 Job Types.
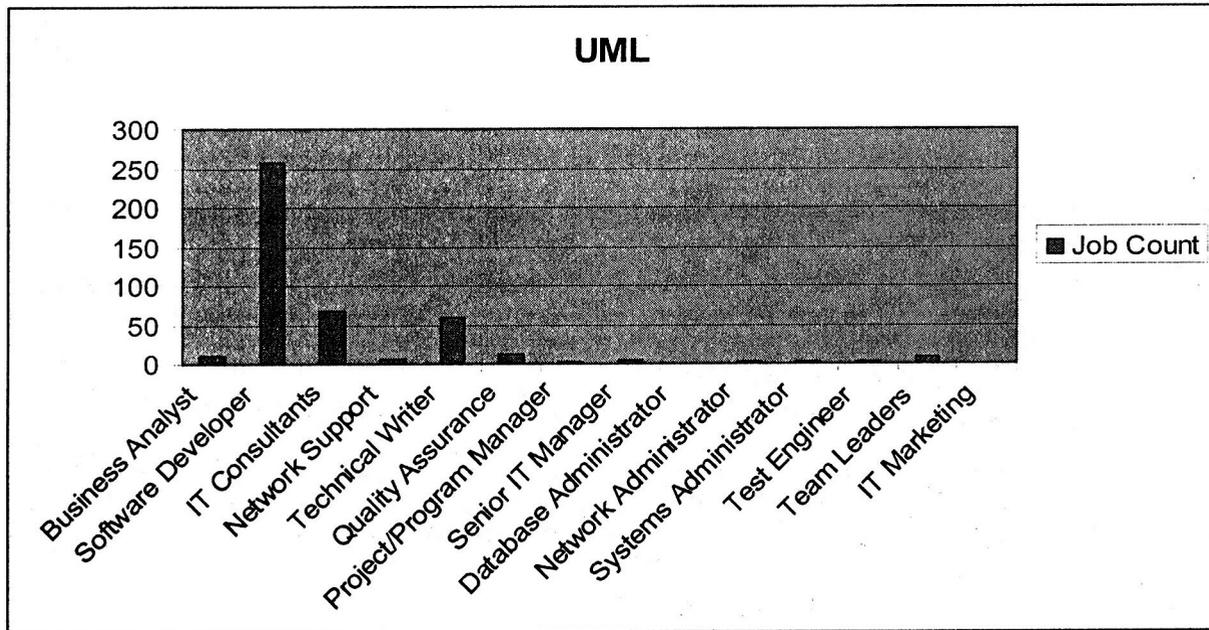
UML

Figure 3(c). Job Counts Containing Keyword 'UML' for the Top 14 Job Types.

Keyword(s) used in the search	Job description of the search results (The keyword used in the search is underscored.)
Data Modeling or Data Model	*Job Title:* Data Warehousing Manager *Required skills:* • Ability to work closely with users to translate requirements into solutions. • Knowledge of process and <u>data modeling</u>, data architectural development methodology, project management, and extensive knowledge of data warehouse and data mart. *Desired skills:* • Large Scale Data Warehousing. • Business Process Analysis. • Category Management Exposure. • Point Of Sale Analysis Exposure. • Extensive experience in a team environment.
ERD(or ERD Tools)	*Job Title:* Database Administrator *Required skills:* • Use of <u>TOAD</u> and SQL*Plus to create and maintain complex PL/SQL stored procedures, functions and packages, including use of cursors, temp tables for intermediate results, user-defined data types, and efficient and maintainable coding practices. • Performance tuning, including interpretation of execution plans, use of hints with the cost-based optimizer, and indexing recommendations . • Advanced application in the creation of complex, dynamic reports, forms and applications, including creation of complex report parameter/prompt windows, use of stored procedures as data sources for reports, and complex report layouts incorporating graphics and multilevel subtotaling *Desired skills:* • Use of Oracle Designer for DDL. • SQL*Loader and the UNIX operating system, including UNIX shell scripting.
UML	*Job Title:* Software Engineer *Required Skills:* • Extensive experience in systems engineering and integration in security programs. • Have strong communication oral and written skills. • Possess demonstrated team leadership skills. *Desired Skills:* • Knowledge of formal modeling (e.g. <u>UML</u>).

	• Experience developing wireless distributed systems. • Experienced in working in a CMM/CMMI environment of Maturity 3 or above. • Understanding of military operations and tactical communications, command, control and military operations.
Data Modeling and UML	*Job Title:* Database Developer *Required Skills:* • Build Package, Store Procedures, Triggers – DBA. • Build and manage tables, synonyms, user accounts, monitor systems. • Object-Oriented Programming -VS.Net (VB, C#, C++), Java. • DirectX/Device programming - Camera, PDA, Hand Geometry, Hardware controller. • Database programming - ADO.NET to Oracle, JDBC to Oracle. • MS Window (2000, XP) API programming, • Service programming – <u>UML</u>, XML/XSD. • Web Services, internet protocol, HTML, Javascript. *Desired Skill:* • Experience in Biometrics as used for identity management. • Familiarity with physical access systems. • Familiarity with systems security concepts.
Data Modeling and ERD	*Job Title:* Database Developer *Required Skills:* • Experience in Information Technology. • Experience creating and implementing physical databases from Logical <u>Data Models</u>. • Experience with data management, versioning of data, backups, etc. • Database Optimization. • Experience in both Oracle and SQL Server DBMS. • Experience with Informix. • Query and Stored Procedure optimization. • Ability to create processes and scripts for change management. • Experience using Database monitoring tools from Oracle, Microsoft or CA. • Experience with <u>Erwin or Similar tool</u>. *Desired Skills:* • Experience with Dimensional Modeling, Star schemas and Data marts. • Experience in the Healthcare industry.
Data Modeling and ERD and UML	*Job Title:* Data Modeler *Required Skills:* • Possess an understanding of <u>data models</u> (conceptual, logical, and physical) or database concepts. • An understanding of modeling concepts (relational, object, star-scheme) . • Understand the mapping between models and standards (<u>UML</u>). • Demonstrated skills in oral and written communications. • This position requires excellent people skills and the ability to interface with customers and other contractor teammates. *Desired Skills:* • Data modeling tool experience in <u>ErWin</u>.

Table 2. Sample Job Descriptions from the Keyword Searched Results.

For UML focused database modeling and design (i.e. Blaha et al. 1997) the key employment opportunity is in 'application', especially in objected-oriented design and not in data modeling or database management.

From both the job counts analysis and content analysis, jobs demand both 'data modeling' and 'ERD' or simply 'ERD' and database analysis tools represent the mainstream definition of data modeling recognized in academic teaching. This includes data models of conceptual, logical and physical analysis and design and database administration practices. When only ERD or its tools like ErWin is mentioned, the skills are focused primarily on database administration. When only UML is referred either with data modeling or by itself, the job focus is mostly on application development and therefore programming skills.

The only place where ERD and UML are both required in data modeling, the mapping between the two methodologies is mentioned. Little indication of using only UML instead of ERD is suggested in the data modeling or in the database administration practice.

4. CONCLUSIONS

Patterns emerged in analyzing the employment skill requirements concerning the methodologies of ERD vs. UML in data modeling. They are summarized in the

following.

- In general information systems, data modeling is one of the required knowledge bases. This explains the large number of keyword searched results in data modeling without specific methodology being referred to.
- Once we focus on data modeling with ERD or other database tools, the more academic data modeling skills including the scopes and concepts of system analysis and design, data and process modeling, data design, relational model, database system, user interface and system support are required.
- Specific methods or approaches of data modeling such as ERD or related tools are not always listed as the required skills and often they are listed as the desired skills. This is in strong contrast to the software engineering or software development discipline, in which the skill requirements of UML are clearly indicated and UML is considered as a standard.
- When ERD is required as a specific methodology, it tends to focus on database design and maintenance and is often mentioned with other tools such as ErWin, Visio or TOAD.
- UML appears to be on the application development for data modeling and is often required as a critical skill.

In summary, we found that there is no simple 'unified' data modeling approach specified in the employment skill requirements. The division in using ERD or UML remains and it is not unlike what is currently found in the commonly used data modeling and database textbooks.

5. ACKNOWLEDGEMENTS

6. REFERENCES

AllFusion® ERwin® data models (2006), retrieved on February 5, 2006 from http://www3.ca.com/solutions/Product.aspx?ID=260

Blaha M.and W. Premerlani (1997), Object-Oriented Modeling and Design for Database Applications, Prentice-Hall.

Gorgone, John T., Gordon B. Davis, Joseph S. Valacich, Heikki Topi, David L. Feinstein, Herbert E. Longenecker, Jr. (2002), IS 2002 Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems, ACM, AIS and AITP.

Hoffer, J.A., M.B. Prescott, F. R. McFadden (2005), Modern Database Management, 7th Edition, Prentice-Hall.

Kroenke D. M. (2005), Database Processing, Fundamentals, Design and Implementation, 8th Edition, Prentice-Hall.

O'Brien , J.O. and G. M. Marakas (2005), Managing Information Systems, 7th edition, McGraw-Hill.

Shelly, Cashman and Rosenblatt (2003), Systems Analysis and Design, 4th edition, Course Technology.

SkillPROOF Inc. (2004), retrieved on August 5, 2005 from http://www.skillproof.com.

Toad™ for Oracle is a database development and administration tool (2006), retrieved on February 5, 2006 from http://www.quest.com/toad/index.asp

Visio 2003 is a diagramming program (2006), retrieved on February 5, 2006 from http://www.microsoft.com/office/visio/prodinfo/overview.mspx

Whitten J. L. and L. D. Bentley (1998), System Analysis and Design Methods, 4th Edition, McGraw-Hill.

AUTHOR BIOGRAPHIES

Hsui-lin L. Winkler is an Associate Professor of Information Systems at the Seidenberg School of Pace University. Prior to joining Pace University, she has many years research experiences in conducting both academia and industry projects using data recorded in natural environments. She also worked in recent years on multi-media database integration and visualization, applied web-enabled technology to design user interface for information navigation and integration. She has taught college courses in database management, object-oriented programming, and system analysis and design. She holds a Ph.D. in Geophysics from California Institute of Technology, Pasadena, CA, and a M.S. in Information System Management from Carnegie Mellon University, Pittsburgh, PA.

Henning Seip is the President and CEO, Co-founder of SkillPROOF Inc. in 2002 with three partners. Mr. Seip also founded The Consultants Network, Inc. (TCN) in 1995, a consulting firm specializing in remote consulting for customers of SAP America. Prior to this he served from 1994 - 1995 as a Vice President and Director of Information Systems at Bantam Doubleday Dell Publishing (today Random House) a 100% owned subsidiary of Bertelsmann AG, Germany. In his capacity Mr. Seip was responsible for the successful restructuring of Bantam's data centers and enterprise software. Mr. Seip received the degree Diplom-Wirtschaftsingenieur in Industrial Engineering from the University of Hamburg (Germany). Mr. Seip has been a speaker at the national Americas' SAP Users Group

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.