AFRL-HE-AZ-TP-2002- 0016

UNITED STATES AIR FORCE RESEARCH LABORATORY

MATCHING JOBS, PEOPLE, AND INSTRUCTIONAL CONTENT: AN INNOVATIVE APPLICATION OF A LATENT SEMANTIC ANALYSIS-BASED TECHNOLOGY

Darrell Laham

Knowledge Analysis Technologies
4940 Pearl East Circle #200
Boulder, CO 80301

Winston Bennett, Jr.

Air Force Research Laboratory
Warfighter Training Research Division
6030 South Kent Street
Mesa, AZ 85212-6061

Thomas K. Landauer

University of Colorado at Boulder
Department of Psychology
Muenzinger D244 345 UCB
Boulder CO 80309-0345

March 2003

AIR FORCE MATERIEL COMMAND
AIR FORCE RESEARCH LABORATORY
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

NOTICES

This research was conducted under the Small Business Innovation Research (SBIR) Program as a Phase I effort. The US Government does not require SBIR Phase I contractors to adhere to any particular format style. In accordance with SBIR guidelines for Phase I efforts, the contractor's report is accepted for publication but not edited.

The views expressed in this report are those of the authors and do not necessarily reflect official views of the US Air Force or the Department Of Defense.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government-related procurement does not in any way obligate the US Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation, or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

The Office of Public Affairs has reviewed this report and it is releasable to the National Technical Information Service where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

WINSTON BENNETT, JR.
Project Scientist

HERBERT H. BELL
Technical Advisor

CURTIS J. PAPKE, Colonel, USAF
Chief, Warfighter Training Research Division

Contract Number: F41624-98-C-5040
Contractor: Knowledge Analysis Technologies

Copies of this report may be obtained at:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, VA 22060-6218
http:// stinet.dtic.mil

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
March 2003

2. REPORT TYPE
Interim

3. DATES COVERED *(From - To)*
May 1998 to February 1999

4. TITLE AND SUBTITLE

Matching Jobs, People, and Instructional Content: An Innovative Application of a Latent Semantic Analysis-Based Technology

5a. CONTRACT NUMBER
F41624-98-C-5040

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
65502F

6. AUTHOR(S)
Darrell Laham
*Winston Bennett, Jr.
Thomas K. Landauer

5d. PROJECT NUMBER
3005

5e. TASK NUMBER
HJ

5f. WORK UNIT NUMBER
8E

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Knowledge Analysis Technologies
4940 Pearl East Circle # 200
Boulder, Co 80301

* Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-AZ-TP-2002-0016

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Air Force Research Laboratory Technical Monitor: Dr. Winston Bennett, Jr., (480) 988-6561 Ext. 297 DSN: 474-6297
A version of this paper was published in Interactive Learning Environment, 8(3), 1-15.

14. ABSTRACT
New latent semantic analysis (LSA)-based agent software helps to identify required job knowledge, determine which members of the workforce have the knowledge, pinpoint needed retraining content, and maximize training and retraining efficiency. The LSA-based technology extracts semantic information about people, occupations, and task-experience contained in natural-text databases. The various kinds of information are all represented in the same way in a common *semantic space*. As a result, the system can match or compare any of these *objects* with any one or more of the others. To demonstrate and evaluate the system, we analyzed tasks and personnel in three Air Force occupations. We measured the similarity of each airman to each task and estimated how well each airman could replace another. We also demonstrated the potential to match knowledge sub-components needed for new systems with ones contained in training materials and with those possessed by individual airmen. It appears that LSA can successfully characterize tasks, occupations, and personnel and measure the overlap in content between instructional courses covering the full range of tasks performed in many different occupations. Such analyses may suggest where training for different occupations might be combined, where training is lacking, and identify components that may not be needed at all. In some instances it may suggest ways in which occupations might be reorganized to increase training efficiency, improve division of labor efficiencies, or redefine specialties to produce personnel capable of a wider set of tasks and easier reassignment.

15. SUBJECT TERMS
FY98 SBIR program; Instructional content; Jobs; Knowledge components; Latent semantic analysis; LSA; People; Phase I SBIR; Training; Training materials

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Ms Liz Casey
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED	23	**19b. TELEPHONE NUMBER** *(include area code)* 480.988.6561 x-188 DSN 474-6188

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

PREFACE

This research was performed as part of a Small Business Innovation Research (SBIR) Phase I contract to Knowledge Analysis Technologies, LLC, for the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division (AFRL/HEA), under Contract F41624-98-C-5040, Work Unit 3005HJ8E. The AFRL/HEA contract monitor was Dr Winston Bennett Jr.

A version of this paper was published in Interactive Learning Environment, Vol 8(3), pp 1-15, in December 2000.

Research and development is currently continuing in SBIR Phase II (Contract No. F41624-99-C-5003). The interactive HeadHunter demonstration system is available on the Internet at http://www.knowledge-technologies.com/HeadHunter/. Please contact Darrell Laham at dlaham@knowledge-technologies.com or Dr. Winston Bennett, Jr., at Winston.bennett@mesa.afmc.af.mil for a password to review the system.

MATCHING JOBS, PEOPLE, AND INSTRUCTIONAL CONTENT: AN INNOVATIVE APPLICATION OF A LATENT SEMANTIC ANALYSIS-BASED TECHNOLOGY

Introduction. Modern organizations are increasingly faced with rapid changes in technology and missions, and need constantly changing mixes of competencies and skill. Assembling personnel with the right knowledge and experience for a task is especially difficult when there are few experts, unfamiliar devices, redefined goals, and short lead times for training and deployment. When too few adequately trained personnel are available for suddenly critical tasks, organizations need the ability (a) to identify existing personnel who could perform the task with the least training, and (b) to create new training courses quickly by assembling components of old ones.

Current solution methods for such problems require large investments of expert labor and are often either unacceptably slow or insufficiently effective. For example, determining whether a particular person's background for a particular assignment requires some time-consuming training component, or allows it to be omitted, would probably necessitate intensive general training research as well as extensive individual questioning. Creating an efficiently individualized training course would be just as difficult. While a well-developed engineering art, in the best case course design, takes many months of specialized analysis and trial.

Thus, more effective methods are desired for characterizing, locating, and training personnel who can optimally perform the set of duties required by any new mission.

This calls for information technologies that can:

(a) represent knowledge and skills,

(b) identify people with all or parts of the knowledge and task experience required by a mission—wherever and in whatever occupation they are currently,

(c) determine precisely what, if any, retraining each person needs in order to perform which new duties,

(d) reduce the effort required to create new training programs, and

(e) minimize the time required for training and retraining.

The objective of this research was to develop and test the practical capability of Latent Semantic Analysis (LSA), in application to these problems. LSA is a machine-learning method for automatically extracting and representing knowledge in massive databases of relevant electronic text (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). It was developed through ten years of basic and applied research supported by Bell Communications Research (now Telcordia), Defense Advanced Research Projects Agency (DARPA), Office of Naval Research (ONR), Army Research Institute (ARI), National Space and Aeronautics Association (NASA), Air Force Research Laboratory (AFRL), the McDonnell Foundation and others. LSA has been extensively validated in both controlled experiments and field tests (Landauer, 1998; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). See Landauer and Psotka (2000) for a fuller explication and for references to publications in which details of method and validation can be found. For demonstrations of the capabilities we describe, see http://www.knowledge-technologies.com/HeadHunter/.

The new personnel data mining application of LSA exploits the explicit and implicit knowledge that already exists in extensive textual computer files of systems documentation, training and test materials, task analyses, and service records. The current research primarily used a small subset of existing Air Force data relevant to 2,121 defined tasks performed in three Air Force occupations by 9,215 airmen. The results showed that the technology could accurately estimate the similarity of each task or occupation to every other task or occupation, measure the degree of match of each airman to every task or occupation, estimate which

airmen could most easily take the place of others, and indicated that LSA has the potential to identify in detail and match the knowledge components required by new systems with those contained in segments of existing training materials, and with the experience of individual airmen. The natural language query design intrinsic to LSA eliminates or alleviates most of the known problems inherent in keyword matching of field-restricted databases.

Description of LSA representations. The biggest advantage of LSA knowledge representation for the present purpose is that different types of data objects (e.g., occupations, job tasks, personnel, training materials) can all exist as vectors within the same semantic space and can therefore be directly compared to each other in meaningful ways. People can easily make holistic judgments of similarity between a task to be performed and a set of people who might be called upon to perform the task. However, the structure in which this information is usually stored in computer files, i.e., in relational databases, has precluded the possibility of automated judgments of this sort.

Traditional database structures are very brittle in that search and retrieval are overly dependent on specific data field choices (e.g., zip code field, job title field) and on exact keyword matching. While in many cases exact matching on highly structured data is desirable (e.g., find the names of all people in zip code 30405), in many other cases the choices can be overly restrictive and/or ambiguous (find all the people who list their job title as "Doctor"). In the latter example, those people who listed their title as Medical Doctor, Physician, Surgeon, General Practice Doctor, and other medical specialties would not match the query and would be inappropriately excluded.

Air Force occupational analysts have successfully employed a *Task Inventory* approach to categorize common tasks within occupational specialties. There are numerous

occupations within the Air Force—the *Airman Classification Structure Chart* details the 190 Specialty Occupations currently active—most of which are non-combat and which have corresponding civilian jobs. Extensive data gathered by the Air Force Occupational Measurement Squadron (AFOMS) is available which describes the various occupations, at least in part, as lists of their relevant task duties. In Air Force Occupational Surveys, personnel are directed to estimate their time spent on these various tasks. These Occupational Surveys served as the primary source of data for the research.

Software programs were developed to quickly create simple text representations from the existing AFOMS databases for different occupations, tasks, and personnel. For example, the software creates an *Airman Object* by gathering selected service status data and the descriptions of tasks each airman specified as having performed. The airman, from LSA's point-of-view, is represented as the list of tasks performed by the airman (and other selected service status items). Additional relevant information, such as a self-description of ability, a resume, or the training courses the person had passed, which would make the description of the person more complete and robust, could easily be added to the personnel representation if available.

The data from three Air Force Specialty Codes (AFSC), Aerospace Mechanics (2A5X1), Physical Therapists (4J0X2), and Medical Services (4N0X1) were used. A semantic space was developed which included 20,000 documents, or object representations. The objects were classified as occupations (full set of tasks for each AFSC; $N = 3$), duty lists (tasks grouped into functional units; $N = 44$), tasks (individual task units; $N = 2,121$) and airmen (task and case record data; $N = 9,215$). Additional randomly selected general knowledge documents ($N = 8,617$) not related to the Air Force data, were included to provide

additional examples of words used in context so that LSA could create a more robust space with a more extensive vocabulary. The total number of objects in the system is 20,000—a very small dataset compared to LSA's capabilities, but enough to do a fair job in estimating statistical regularities for the three occupations.

By adding less than 9,000 general reading documents, the semantic representations of more than 50,000 additional words were included in the space. This additional representational power allows user synonyms to match task analysis keywords. For example, the space would consider *tire* to be a synonym for *wheel*, or *physician* a synonym for *doctor* even if *tire* and *physician* were never used in the task analysis data.

Figure 1 shows a two-dimensional (2-D) representation of two LSA objects, a Job and a Candidate. In actual LSA representations, it requires 100-500 orthogonal dimensions to characterize an object—this 2-D representation is for illustrative purposes only. Unlike traditional factor analysis, where the dimensions have been interpreted and named, LSA dimensions are not *in and of themselves* meaningful. Both objects are seen as points in the Semantic Space having some score on the X dimension and an independent score on the Y dimension. In most of the work reported in this paper, each of the objects has scores for 300 orthogonal dimensions, thus each object is represented as a vector of 300 numbers, rather than a vector of two numbers. In actual LSA, the meaning of an object is determined when the full set of dimension scores is used in the comparison of the object to other objects. In most cases LSA uses the cosine of the angle between objects as the measure of similarity.

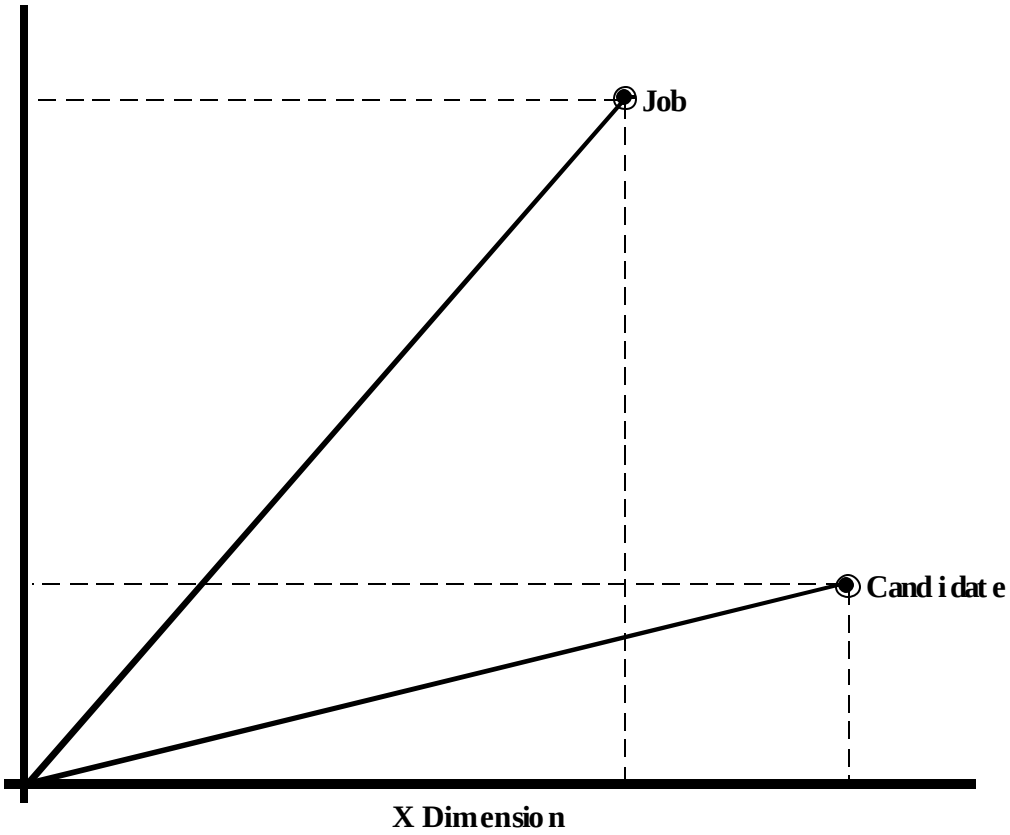

Figure 1. A 2-D representation of two LSA objects. In this representation, it is seen how quite different objects, here a description of a Job and a description of a Candidate for the Job, are cast in the same semantic space in LSA. Both objects are seen as having some score on both dimensions. In actual LSA Semantic Spaces, objects are represented using from 100-500 dimension scores.

In Figure 2 the capacity to make similarity comparisons between different data objects is illustrated. To determine which of two candidates is best suited for a job, the LSA system would use the smaller of the two angles 1 and 2. Cosines range between 1 and -1 for all possible angles—an angle of 0 degrees has a cosine of 1 (the vectors lie on top of each other), an angle of 90 degrees has a cosine of 0 and an angle of 180 degrees has a cosine of -1. The smaller the angle, the higher the cosine, and the more similar the two objects are considered. In the Figure 2 case, Candidate 1 is a better choice for the Job.

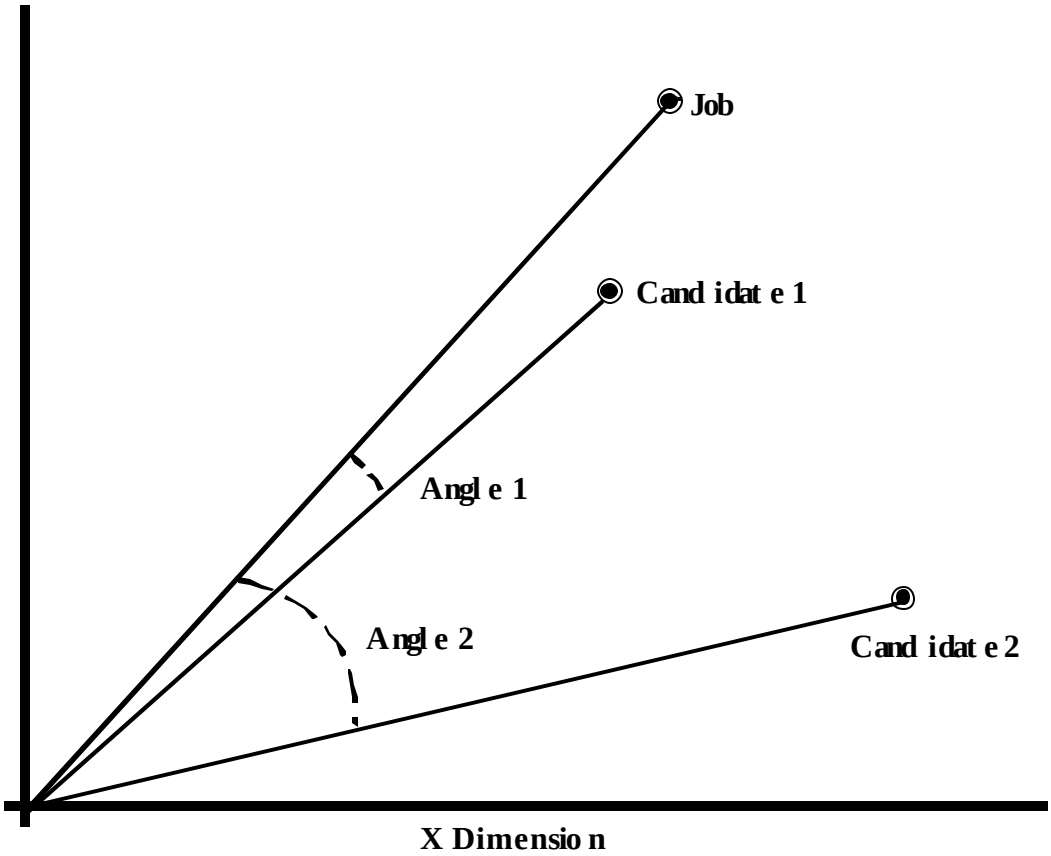

Figure 2. Two candidates for a job. In LSA any point in space can be compared to any other point, thus disparate types of objects can be transparently compared. Using the cosine of the angle between vectors, we can compare both Candidates to the Job. In this 2-D illustration, it can easily be seen that Candidate 1 is more like the Job than is Candidate 2. All else being equal, LSA would suggest that Candidate 1 is the better match for the Job.

In Figures 3 and 4, the scenario of selecting the best training for a Candidate seeking a particular Job is demonstrated. In Figure 3 the vectors for the Job and the Candidate are displayed as before, along with the vectors for two training Courses (A and B) that have been designated as appropriate for the Job. In Figure 4 it is seen that while both Courses contain knowledge important for the Job, the appropriate choice (Course A) as determined through LSA would bring the Candidate much closer to the Job vector than the alternative choice.

Given a larger set of possibly related courses to choose from, the situation could be somewhat more complicated. For example, given three courses, A, B, and C containing knowledge a particular airman needed but lacked, A might be at too advanced a level for the

individual to understand given his or her previous knowledge, B might overlap too much with what the individual already knew to be of much value, and C might be "just right", containing just the new knowledge that the person is ready to learn. LSA methods capable of picking the optimum text under these circumstances have also been developed and empirically demonstrated (Wolfe et al., 1998; Rehder et al., 1998)

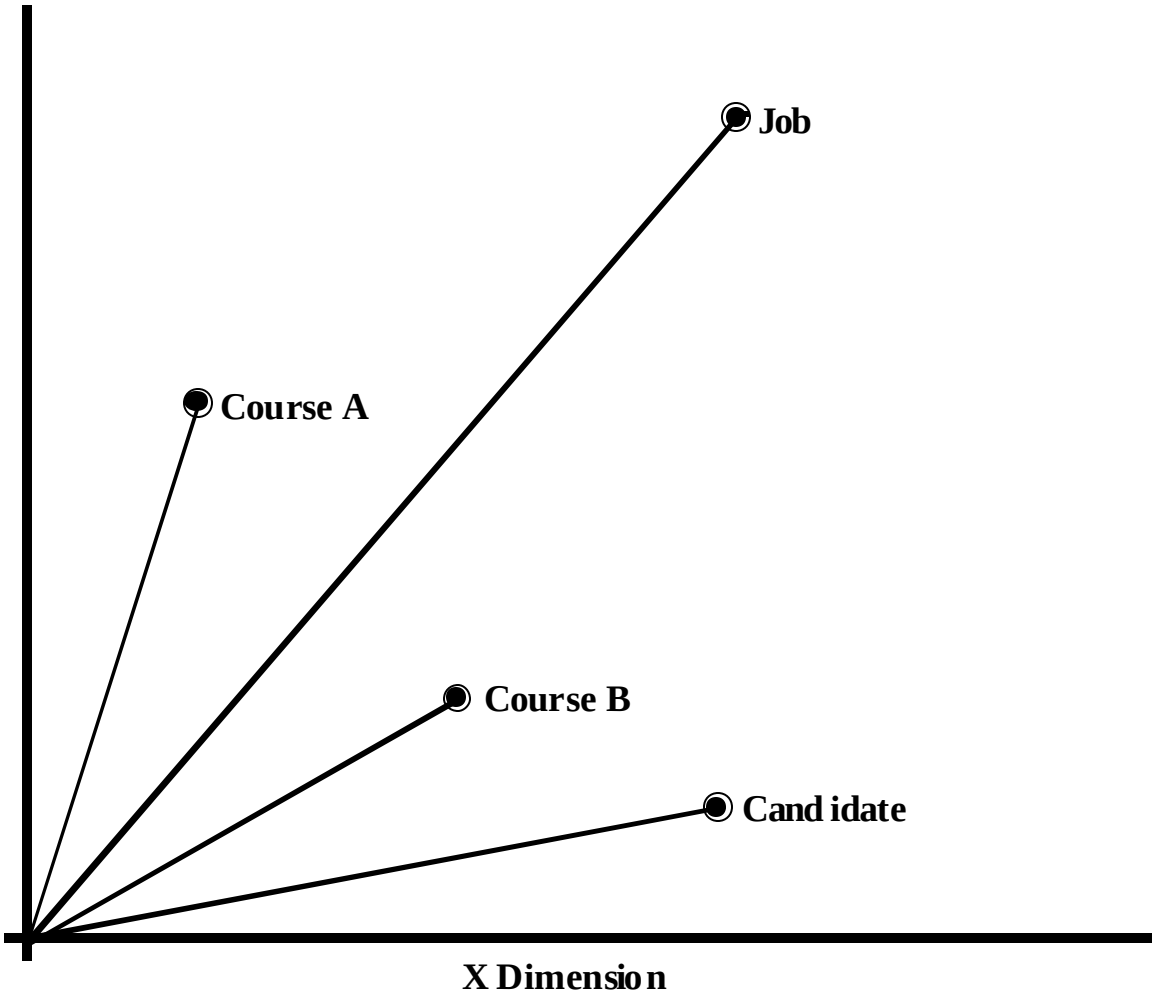

Figure 3. A job, a candidate, and two training courses for the job. Both Course A and Course B have been identified as appropriate training for the Job, and both are courses that the candidate has not yet taken. Which Course is most appropriate for the Candidate to take?

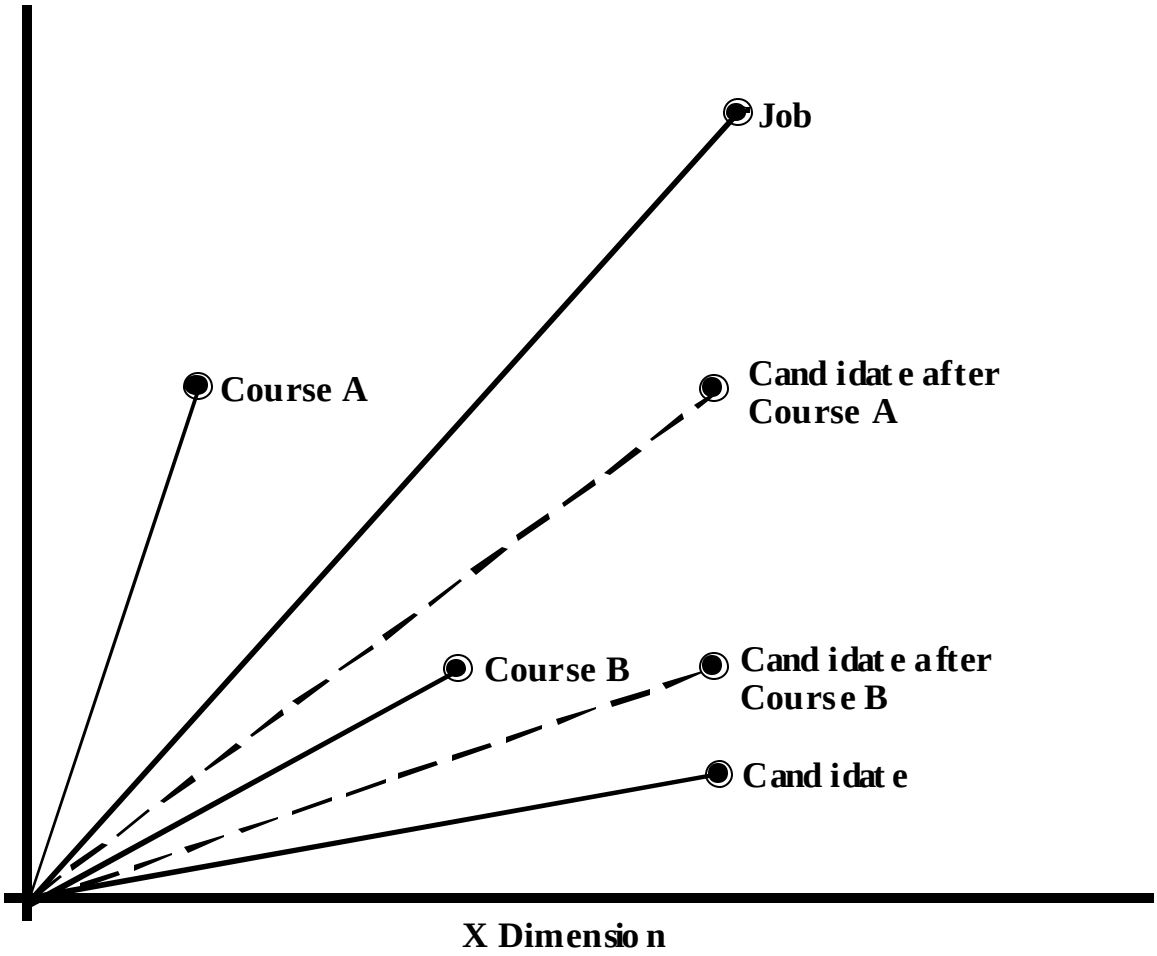

Figure 4. The candidate after taking either training course. Should the Candidate have been assigned Course B, he would have moved slightly closer to the job, however, much of the content of the course was information which the Candidate already possessed. After taking course A, the Candidate became much more prepared for the Job. The primary content of Course A was the information that the Candidate previously lacked. The averaged vectors for *Candidate after Training* are overly simplified in this illustration—the knowledge represented as the course in actual LSA would be distributed over hundreds of dimensions. Each dimension does not correspond to a simple, nameable concept.

Results from Modeling Studies. Several experiments were conducted over the course of this research to validate the quality of the representations within the developed semantic spaces for these purposes.

Course content overlap. We conducted a study on the similarities of content between a set of occupational specialty training courses based on the content of their final exam items.

We were provided with text of the multiple-choice question and correct alternative answer for each of 100 items for each of 95 specialty courses (AFSC). For security, the content was encrypted before being sent by replacing each unique word wherever it appeared with a previously assigned string of random digits. From this we constructed a matrix of ~9,500 test items by ~20,000 unique word-types and performed LSA at 300 dimensions to obtain a vector representing each test item.

To test how well LSA had captured the semantic similarities of the domain, we computed the similarity, measured by cosine, of each item with that of 100 randomly chosen items from the same test (*Within*) and with 100 randomly chosen items from all other tests (*Between*). Item pairs from the same test had higher average cosines over all tests (n = 9,500, t = 35.2, $p < .0001$), significantly so in 82 of 95 of the individual cases (n = 100, $p < .05$). See Figure 5.

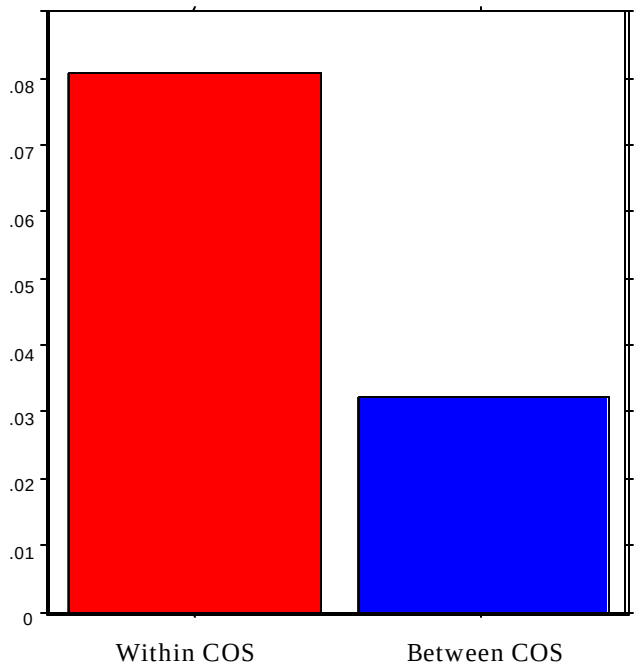

Figure 5. Average cosines within and between course test items

We also created an LSA vector for each course by finding the centroid or average, of the 100 test items for each course. A subsequent hierarchical clustering of the 95 centroid vectors for courses, done blind, produced groupings later judged to be intuitively meaningful and useful by knowledgeable Air Force training researchers. The cluster analysis grouped courses for similar specialties together and revealed content redundancies between the courses that were useful in subsequent course restructuring.

Comparison of Airman Similarity. In an analysis of the semantic space developed from the AFOMS data on 9,215 airmen, the system's estimates of similarity for airmen within a specialty were compared to the judgments of similarity between specialties. The average LSA similarities were calculated between 3,000 randomly sampled pairs of airmen either within an occupation, or between occupations. Each airman was compared with a random sample of others both within their own AFSC and between the alternative AFSC. The LSA similarities of airmen within an AFSC were significantly greater ($p < .0001$) than the similarities between airmen in different AFSCs (see Table 1 and Figure 6). The specialties that were more diverse showed less similarity for the airmen within the specialty than did the specialty that was more uniform. Subject-matter experts informally concurred with the judgments of similarity for a small random sampling of the 42,453,505 independent similarity judgments available with a sample size of 9,215 airmen.

Table 1. Similarities for occupational specialties

2A5X1 within	0.60	4J0X2 to 4N0X1	0.23
4J0X2 within	0.92	2A5X1 to 4J0X2	0.12
4N0X1within	0.60	2A5X1 to 4N0X1	0.21
AVERAGE WITHIN	**0.71**	**AVERAGE BETWEEN**	**0.19**

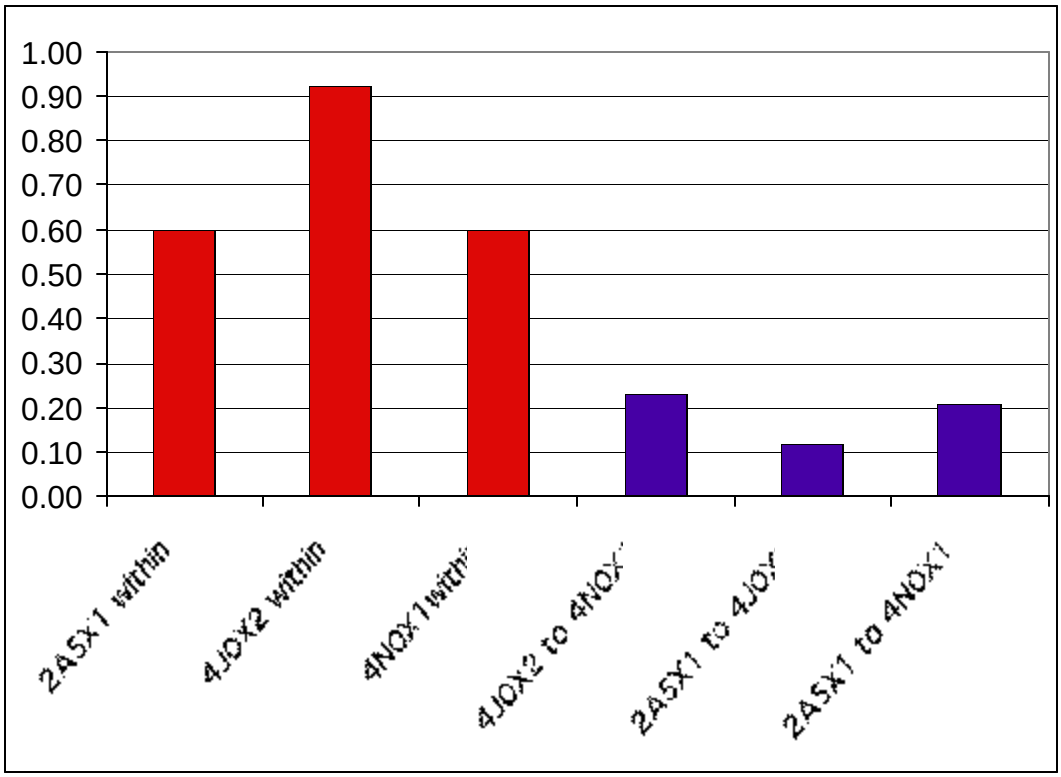

Figure 6. Mean cosine of personnel comparisons within and between occupational specialties

Clustering of Duty List Items. Another experiment looked at how the representations for the three Occupations and the 44 Duty Lists (tasks grouped into functional units) clustered in the semantic space. The analysis shows the cluster distribution of shared and unique duty lists among the occupations. The within specialty duty lists clustered tightly around their respective occupation, while those duties that are shared across occupations, such as training, management, and supervisory tasks, clustered together and distinct from the occupations regardless of originating specialty and differences in specific task representations.

Potential Applications of the Method.

Job Placement or Occupation Assignment. Practical applications to job assignment were most directly illustrated by the research just reported. The simplest case is direct replacement of one airman with another. For this, a query takes the form of the to-be-replaced airman's identification number, and the *k* most similar airmen known to the system

(potentially all those in the Air Force plus others where relevant) are returned and listed in terms of their overall task-experience pattern—the closeness of their points in the joint semantic space representing tasks, occupations, and airmen. Their complete service records can then be displayed and compared. If it is desired to add a new member to a work group, the descriptions of those tasks that are most in need of additional help can be entered as the query and the system will list in order those airmen whose total experience is most like the new job requirements. Note that in performing this match, LSA goes beyond simply counting the number of tasks in common between the wanted list and the service record, instead factoring in previous experience (and, later, training) in occupations and tasks that are similar but not identical to those in need of performance. Thus, it would be quite possible, in the absence of any airman who had done any of the prescribed tasks, to nevertheless find one or more candidates who had done similar work, the estimate of similarity having been automatically induced by LSA from the entire corpus of data without human intervention.

The technique could be used to add people to perform new jobs, by adding to the query a free-form description of the tasks involved. Because LSA captures semantic and conceptual similarity of verbal expressions, it will correctly match *ad hoc* task descriptions with official task definitions and job descriptions. The system can also form a representation of the overall mix of tasks required by a group by combining representations of the knowledge possessed by all its present members. In case of downsizing, the system would make it possible to find a set of personnel to transfer out of a group that would either leave it most like its previous composition, or desirably modified, again without relying on a crude counting operation or intuition.

The opportunity and manner of application for selecting airmen for missions, for example, expeditionary war-fighting missions with unique challenges, is relatively straightforward. Given a careful verbal description of the mission, including all the tasks to be performed, the equipment, weapons, devices, procedures, numbers of airmen needed in each role, and perhaps even factors such as locale, terrain, and likely weather and other challenges, the LSA matching technique would rank airmen for suitability to each task on the basis of the totality of their previous task and occupational experience, along with, if available, relevant (as determined by LSA) test scores and performance ratings.

Curriculum Overlap Analyses. The Air Force (like other military and civilian organizations) offers hundreds of specialty training courses, many of which overlap substantially in content, many of which may contain content no longer relevant to tasks currently in demand, and some which are missing content made desirable by changes in technology, missions, or staffing. In many cases it would be desirable to combine, condense, or modify courses. Teaching unnecessary numbers of courses or redundant components in multiple courses is expensive in instructional staff and facilities, and even more expensive in wasted student time and resources. Teaching material that is sub-optimally matched to work requirements, either by being superfluous, redundant, or by failing to equip airmen with the best skill sets for all the tasks it would be desirable for them to be able to perform, is probably even more expensive in the long run.

To rationalize the content and organization of content for multiple training programs, a method is needed by which the overlap in course content can be easily assessed. Presently such analyses are performed, if at all, by highly labor-intensive efforts by subject-matter experts and training specialists. We have already demonstrated that LSA can do this kind of

analysis automatically to a quite useful degree. Our studies were based on analysis only of course examination items, but appeared to give a great deal of useful information about course overlap.

LSA can also measure the overlap between course content and the full range of tasks performed in many different occupations. Information from such analyses will suggest where the training needed for different occupations overlaps and might be combined, where training is lacking, point to components that may not actually be needed at all, and, in some instances, suggest ways in which occupations might be restructured to increase training efficiency. LSA methods will not solve these problems completely, but we believe they can offer highly useful information for planners that is currently unobtainable or prohibitively expensive.

Just-in-timeTraining Materials. In brief, the way in which we envision that LSA would be employed in helping to rapidly create new targeted training programs might be as follows. The component knowledge needed and tasks to be performed for a new device, system, or procedure would be carefully described by designers and relevant subject-matter experts. LSA would determine the degree of match of each component to a wide range of tasks performed in the Air Force and to every paragraph in every possibly relevant training or operations manual. Tasks and paragraphs would be ranked by estimated relevance to the new system, and the LSA similarity of each paragraph to each task determined. In the quickest and dirtiest version, a custom retraining document for each candidate could initially be compiled from paragraphs highly relevant to the new system that are not highly similar to tasks the candidate has previously performed. In the case of urgent need for a small number of trainees, a subject-matter or training expert could then edit each version. In case of need for large numbers and more available time, the collection of paragraphs could be crafted into

a simple computer-based training program with branching to permit trainees to skip parts they already know.

HeadHunter—A WWW-based Intelligent Search Agent. The HeadHunter software represents an initial demonstration of a usable World-Wide Web-based Intelligent Search Agent based on the LSA technology. Currently the system has knowledge of the three AFSCs used in the research, however, even with its limited knowledge base, it demonstrates the necessary capabilities to match mission and job requirement statements with military personnel and training data. By measuring semantic similarity of training materials and tests, it promises to facilitate combining occupations based on core competencies and similar work activities. It also promises to help identify individuals qualified for work activities for which no current occupation exists.

An organization that has acquired a new or revised system could develop detailed descriptions of the activities required to operate or to maintain the system, based on system requirements documents, operations manuals, or provided by subject-matter experts. Given such descriptions, and assuming an increase in its knowledge, HeadHunter could automatically identify current jobs on fielded systems that are similar in component work activities and in their requirements for training. It would also identify similar paragraphs in existing course materials and rank them by probable relevance to work with the new system.

New occupations could be structured around these activities and new sets of training materials assembled, at least in major part, from subsets of existing material. In addition, individuals who work in jobs that use subsets of the competencies and experience required can be identified. This might permit the immediate employment of appropriate personnel or their more rapid and effective re-training for work in support of new systems.

In the occupational domain, this effort may ultimately produce a cost-effective capability to systematically mine occupational personnel and training databases to develop new job and training structures to support a variety of requirements. This capability would help employers identify critical characteristics and competencies associated with work activities and then to identify individuals who have the requisite experience and competencies to perform the identified work activities.

Conclusions. The preliminary investigations using three Air Force specialties demonstrate the potential of the LSA-based system for personnel and occupational analyses. The ability to create holistic LSA representations from pre-existing structured databases is an important capability for development of the technology for practical applications. The importance of using these techniques with large databases, e.g., for the hundreds of thousands of personnel in branches of the military, or even the thousands of employees in large corporations, should be emphasized. The LSA tools allow for analyses that have been heretofore impossible because of the size and complexity of the involved data.

REFERENCES

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing By Latent Semantic Analysis. Journal of the American Society For Information Science, 41(6), 391-407.

Landauer, T. K. (1998). Learning and representing verbal meaning: The Latent Semantic Analysis theory. Current Directions in Psychological Science, 7, 161-164.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104, 211-240.

Landauer, T. K, Foltz, P. W. & Laham, D. (1998) An Introduction to Latent Semantic Analysis. Discourse Processes, 25(2&3), 259-284.

Landauer, T. K., Laham, D., & Foltz, P. W., (1998). Learning human-like knowledge by Singular Value Decomposition: A progress report. In M. I. Jordan, M. J. Kearns & S. A. Solla (Eds.). Advances in Neural Information Processing Systems 10, 45-51. Cambridge: MIT Press.

Landauer, T. K., Laham, D., Rehder, B. & Schreiner, M.E. (1997). How well can passage meaning be derived without using word order: A comparison of Latent Semantic Analysis and humans. In M. G. Shafto & P. Langley (Eds.), Proceedings of the 19th annual meeting of the Cognitive Science Society (pp. 412-417). Mahwah, NJ: Erlbaum.

Landauer, T.K., & Psotka (2000). Simulating text understanding for educational applications with Latent Semantic Analysis: Introduction to LSA. Interactive Learning Environment, 8(2).

Rehder, B., Schreiner, M. E., Wolfe, M. B, Laham, Landauer, T. K, & Kintsch, W. (1998). Using Latent Semantic Analysis to assess knowledge: Some technical considerations. Discourse Processes, 25, 309-336.

Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and text by Latent Semantic Analysis. Discourse Processes, 25(2&3), 309-336.