

Gathering Requirements for a Grid-based Automatic Marking System

Debra Haley, Anne de Roeck, Bashar Nuseibeh, Marian Petre, Pete Thomas

▶ **To cite this version:**

Debra Haley, Anne de Roeck, Bashar Nuseibeh, Marian Petre, Pete Thomas. Gathering Requirements for a Grid-based Automatic Marking System. 1st International ELeGI Conference on Advanced Technology for Enhanced Learning, 2005, Vico Equense (Naples), Italy. pp.8. hal-00190560

HAL Id: hal-00190560

https://telearn.hal.science/hal-00190560

Submitted on 23 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gathering Requirements for a Grid-based Automatic Marking System

Debra Trusso Haley, Anne De Roeck, Bashar Nuseibeh, Marian Petre, Pete Thomas

Computing Research Centre, Department of Computing
The Open University, Walton Hall, Milton Keynes, UK MK7 6AA

D.T.Haley; A.DeRoeck; B.Nuseibeh; M.Petre; P.G.Thomas [at] open.ac.uk

This paper reports on our experiences using a Creative Requirements [1] workshop approach to elicit requirements for a Grid-based automatic marking system. The research was conducted for ELeGI, an EU funded project whose goal is to provide a European Learning Grid infrastructure to promote a learning paradigm shift from a teacher-centred approach to a learner-centred approach. The automatic marking system uses Latent Semantic Analysis (LSA) to assess the meaning of essays written by computer science students. We foresee the marking system to be a service offered by the Learning Grid. The Creative Requirements Workshop used eight creativity triggers and testimony from an expert witness to elicit creative requirements from the participants. The participants in the workshop produced over 200 requirements in about two hours.

1. INTRODUCTION

The study described in this paper was conducted by researchers at the Open University for a European project called ELeGI (European Learning Grid Infrastructure). ELeGI is a large and diverse project; the portion of the project we are concerned with is an automatic marking system to be offered as a Grid service to ELeGI users. The marking system will grade essays written by computer science students. Educators prefer assigning essays instead of multiple choice questions to more deeply assess a learner's knowledge. Unfortunately, marking essays is a very time-consuming task. Another drawback is the inherent subjectivity of human-marked essays. A computer-based marking system will overcome each of these drawbacks.

Before building the marking system, we need to compile a complete set of detailed requirements. The Open University is a natural place to find requirements since it has specialised in distance education for 30 years. Its instructors and students would benefit from a well-designed, speedy, accurate, always-available marking system.

We conducted a Creative Requirements Workshop in October 2004 to gather requirements for an LSA-based automatic marking system to be developed for ELeGI.

Section 2 introduces the ELeGI project. Section 3 gives examples of some of the many existing applications that use LSA and explains the mathematics behind the LSA algorithm. Section 4 summarises the information presented by Robertson and Robertson at their Creative Workshop tutorial [1]. Section 5 discusses the workshop presented at the Open University and lists some of the over 200 requirements that were elicited. Section 6 concludes with lessons learned and suggestions for future work.

2. ELEGI

"The European Learning Grid Infrastructure (ELeGI) project has the ambitious goal to develop software technologies for effective human learning. With the ELeGI project we will promote and support a learning

paradigm shift. A new paradigm focused on knowledge construction using experiential based and collaborative learning approaches in a contextualised, personalised and ubiquitous way will replace the current information transfer paradigm focused on content and on the key authoritative figure of the teacher who provides information."[1]

ELeGI is a large, four-year project with 23 partners from nine European countries. The Computing Department of the Open University will produce an automatic essay grading system for computer science students using an algorithm based on Latent Semantic Analysis (LSA). The Creative Requirements Workshop reported on in this paper was held on behalf of ELeGI to elicit requirements from potential users of the automatic marking system.

3. LSA

LSA is a statistical natural language processing technique used to infer meaning from a passage of text. It was invented in the late 1980s to overcome some of the difficulties inherent to the keyword approach for information retrieval [2]. Expanding from its original use for information retrieval, Landauer and Dumais claimed in 1997 [3] that LSA is a model of word acquisition and human cognition. They found that an LSA-trained system scored as high as the average foreign student taking a TOEFL test (Test of English as a Foreign Language) [4]. Section 3.1 lists some existing LSA-based applications. Section 3.2 explains how LSA works.

3.1 Existing LSA Applications

Since the 1990s, researchers have been using LSA to analyze word meaning in various innovative applications. Wolfe, et al. [5] have achieved good results using LSA to match readers with texts of the appropriate level. If a text is too easy, a learner doesn't learn anything new; if a text is too hard, it is incomprehensible thus preventing the learner from learning anything. Foltz, et al. [6] refer to the Goldilocks principle in using texts at just the right difficulty level – slightly beyond the learner's ability and knowledge.
Some interesting research is being conducted into medical uses of LSA. Wu et. al. [7] are using LSA to classify protein sequences. Skoyles [8] is using the Landauer-Dumais theory of language acquisition [3] to investigate whether autism results from a failure in an individual's ability to create meaning by an indirect process - the induction modelled by LSA. Campbell and Pennebaker [9] are using LSA to demonstrate linkages between writing about traumatic events and improving health.

Much work is being done in the area of using LSA to grade essays automatically and to provide content-based feedback. One of the great advantages of automatic assessment of essays is its ability to provide helpful, immediate feedback to the student without burdening the teacher. This application is particularly suited to distance education, where opportunities for one-on-one tutoring are infrequent or non-existent [10].
Existing systems include Apex [11] AutoTutor [12], Intelligent Essay Assessor [6], Select-a-Kibitzer [13], and Summary Street [10] [14]. They differ in details of audience addressed, subject domain, and advanced training required by the system [13]. They are similar in that they are LSA-based, web-based, and provide scaffolding, feedback, and unlimited practice opportunities without increasing a teacher's workload [10]. See [13] for an excellent analysis of these systems.

3.2 How LSA works

LSA is a specialisation of the vector space model. It induces word meaning by examining co-occurrence data over a large corpus of text. The major difference between the vector space model and LSA is that LSA reduces the vector space to a smaller size thereby reducing the noise caused by chance usage and individual style. [3]

To use LSA, researchers amass a suitable corpus of text. They create a term-by-document matrix where the columns are documents and the rows are terms [15]. A term is a subdivision of a document; it can be a word, phrase, or some other unit. A document can be a sentence, a paragraph, a textbook, or some other unit. In other words, documents contain terms. The elements of the matrix are weighted word counts of how

[1] http://www.elegi.org

many times each term appears in each document. More formally, each element, a_{ij} in an i x j matrix is the weighted count of term i in document j.

LSA decomposes the matrix into three matrices using Singular Value Decomposition (SVD), a generalisation of factor analysis. Deerwester et. al. [15] describe the process as follows.

Let t = the number of terms, or rows
d = the number of documents, or columns
X = a t by d matrix

Then, after applying SVD, X = TSD, where

m = the number of dimensions, m <= min(t,d)
T = a t by m matrix
S = an m by m diagonal matrix, i.e., only diagonal entries have non-zero values
D = an m by d matrix

LSA reduces S, the diagonal matrix created by SVD, to an appropriate number of dimensions k, where k << m, resulting in S'. The product of TS'D is the least-squares best fit to X, the original matrix [15].

The literature often describes LSA as analyzing co-occurring terms. Landauer and Dumais [3] argue it does more and explain that the new matrix captures the "latent transitivity relations" among the terms. Terms not appearing in an original document are represented in the new matrix as if they actually were in the original document [3]. LSA's ability to induce transitive meanings is considered especially important given that Furnas et. al. [16] report fewer than 20% of paired individuals will use the same term to refer to the same common concept.

LSA exploits what can be named the transitive property of semantic relationships: If A→B and B→C, then A→C (where → stands for is semantically related to). However, the similarity to the transitive property of equality is not perfect. Two words widely separated in the transitivity chain can have a weaker relationship than closer words. For example, LSA might find that copy → duplicate → double → twin → sibling. Copy and duplicate are much closer semantically than copy and sibling.

Finding the correct number of dimensions for the new matrix created by SVD is critical; if it is too small, the structure of the data is not captured. Conversely, if it is too large, sampling error and unimportant details remain, e.g., grammatical variants [15], [14]. Empirical work involving very large corpora shows the correct number of dimensions to be about 300 [3], [14].

Creating the matrices using SVD and reducing the number of dimensions, often referred to as training the system, requires a lot of computing power; it can take hours or days to complete the processing [13]. Fortunately, once the training is complete, it takes just seconds for LSA to evaluate a text sample [13].
An automatic marker using LSA is essentially a text-categorizer [17]. Training data comprises both a general corpus and a specific corpus of human-graded essays [18]. A student essay to be marked is evaluated by finding the mark attached to the essay in the corpus that is the closest match [19].

4. CREATIVE REQUIREMENTS WORKSHOPS

We conducted a workshop to elicit requirements for a Grid-based automatic marking system. The workshop techniques described in this paper are based on the Robertson and Robertson tutorial [1] presented at the IEEE Joint International Requirements Engineering Conference in Kyoto on 6 September 2004. The ideas that follow are not necessarily unique to the Robertsons but are a summary gleaned from their research and experience in the field of requirements engineering.

4.1 Creativity triggers

Since customers do not always know what they want until they see it, requirements engineers need to be creative. But being creative isn't easy. The Robertsons offer eight triggers to spur creative thinking during a requirements gathering workshop.

- service
- speed
- information
- participation
- connectivity
- convenience
- origin of the business event
- technology

Each of these triggers is used to encourage the workshop participants to invent or improve a product. The workshop participants are asked to consider one trigger at a time and to brainstorm as many requirements as possible. The triggers are explained below.

Service can provide extra value to a product. It can make the difference in persuading potential customers to choose *your* product instead of your competitors' products. This trigger suggests that workshop participants should focus on how to improve the service associated with their product.

In this fast-paced world, customers appreciate speedy access to a product, whether it be a package delivered in the mail, pricing information, or ticket availability. Participants aim to identify services or products that can be delivered to the customer faster as they focus on the speed trigger.

Information and choices provide added value to a product or service. For example, an agency that provides online seating charts to theatre-goers might increase bookings. This trigger encourages participants to imagine types of information or choices that their customers would appreciate.

The participation trigger suggests that customers like to participate in their purchases. For example, mailing services allow customers to track the status of their packages, online book sellers encourage their customers to write book reviews, and travel companies offer customers the ability to book their own reservations. Workshop participants are challenged to find ways for their potential customers to participate in the purchasing or managing of a service or product.

Providing customers with loyalty cards, frequent flyer programs, and newsletters are ways to make them feel connected. The connectivity trigger states that you can keep your customers by making them feel connected to your business. This trigger asks participants to find ways to connect to their customers.

More convenient products or services are very much appreciated by customers. People are often willing to pay more (because of shipping and handling charges) for online shopping because it is more convenient. Participants respond to the convenience trigger by finding ways to make the lives of their customers easier.

Businesses can gain customers when they consider the origin of the business event and find ways to extend the product boundary. New housing builders do this when they realize that their customers are doing more than buying houses – they are creating a home. People who buy houses also need to buy lighting fixtures, landscape their gardens, and familiarise themselves with the new neighbourhood. Builders can increase their profit by providing extra products, referrals to local businesses and relocation assistance. Workshop participants can create added value by finding ways to extend the boundary of their product or service.

The technology trigger is a familiar technique for many engineers. They invent a technology and then search for a clever way to use it. Workshop participants are asked to find innovative ways of using a new or familiar piece of technology.

Each of these creativity triggers can be used to stimulate ideas by focusing on certain areas one at a time. It is important for the workshop participants to work in groups to allow them to feed off the ideas of each other.

4.2 Expert testimony from another domain

The second technique offered by Robertson and Robertson is working with experts [1]. They suggested that the knowledge and experience of experts in other domains can be used to spark creative requirements through the use of "analogical mappings between actions, agents, objects, requirements, design features, constraints, etc" [1]. Table 1 is taken from [1]; it lists three types of creativity and some professionals whose testimony could be fruitful for requirements engineers.

Table 1: Working with experts

For exploratory creativity:
Cartoonists: discover new angles on subject matter
Fashion designers: understanding future trends
Musicians: creating new testable artefacts
For combinatorial creativity:
Fusion chefs: willingness to combine ideas with reason
DJs: combining different music styles according to rules
Cocktail bartenders: understanding what to mix
For transformational creativity:
Software/systems engineers: transform specifications
Science writers: transform something to understand it
Graphic artists: produce graphics from concepts

5. DISCUSSION AND RESULTS

The Creative Requirements workshop reported on in this paper took place at the Open University in October 2004. The participants were 13 academics ranging from PhD students to professors. The participants listened to a presentation on the eight creative triggers. They were then divided into groups of three or four and asked to generate requirements for the LSA-based automatic marking system described in Sections 2 and 3. They were given sheets of paper with each sheet containing one of the creative triggers. A total of 157 requirements was elicited in about 15 minutes. Space limitations do not permit a complete listing of the results, but several of the more common or interesting requirements for each trigger are shown below. Note that the requirements are given as supplied by the workshop participants. Thus, many need to be clarified and refined. Some of them specify a solution instead of a requirement; for example a participant suggested web cams for monitoring when a cleaner requirement would be "provide a way for the examiners to monitor the test-takers".

service

- provide information on how to improve the answer
- provide pointers to other sources
- detect plagiarism
- bigger marks for visually impaired students
- provide marks via email or letter
- provide a timer to help students pace themselves
- spelling and grammar correction
- make grading rubrics explicit
- highlight gaps in learning through feedback to examiner – individually or as a group

speed

- more efficient for professor and school
- quick feedback to students
- quick processing of norms
- quick distribution of statistics

information

- rationale for the marking
- statistics for the class
- plagiarism detection
- pointers for self-improvement

- information on own and group performance
- health diagnosis – stress symptoms
- how close the student's answer is to the perfect answer

participation

- training course for markers
- ask learners what style comments they want
- get students to self-evaluate, co-evaluate
- use web cams to let examiner watch test being taken

connectivity

- read to develop understanding
- Bluetooth exam papers
- SMS marking
- SMS submission
- give marking scheme before assessment
- get back to the student with assessor's comments and allow student appeal
- available 24/7 – students can modify at the last minute

convenience

- take exam when wanted
- career planning service
- electronic appeal service
- take an exam anywhere, e.g. on the bus
- get feedback on PDA/mobile phone

origin of the business event

- provide sample exams for "learning outcome" to attract students
- preparation of future exams to cover learning holes
- event – take exam; may be thinking about next course – so offer new courses
- may be stressed – online stress advice

technology

- handwriting recognition
- voice recognition
- distance – not tethered
- use a PDA/phone
- online exams – with question timers for pacing
- support webcams for monitoring
- digital signature for authenticity of exam answers
- provide AI to suggest better phrasing of sentences
- must handle network traffic – because usually everybody tends to submit in the last minute

After the participants completed generating requirements using creative triggers, they listened to a presentation on how to use expert testimony to create requirements from analogical mappings. For our workshop, "Andy the auto mechanic" gave a ten minute talk about the job of being an auto mechanic. Andy reported that his work required three phases – information gathering, problem determination, and problem correction.

Information gathering can be tricky since what the customer says happens can not always be reproduced. For example, the customer might say the car goes "kawhirr kawhirr" when he starts it but Andy hears nothing in the starter system that makes a "kawhirr" noise. In fact, Andy senses an unusual vibration when the starter motor engages.

In order to determine the problem, Andy uses several diagnostic tests and instruments including compression tests, vacuum tests, oil pressure, electrical generation, and timing (cam, spark, choke, etc). Another useful procedure is to study the fault trees supplied by manufacturer.

Once Andy diagnoses the problem, he can begin to fix it. He uses special tools for such operations as removing a motor, aligning a clutch and transmission, and aligning brakes. He refers to assembly and repair manuals provided by the manufacturer.

Our participants produced 46 requirements in about 15 minutes after listening to the expert and trying to form analogies between the work of an auto mechanic and an automatic marking system. Some of the more common or interesting ideas are listed below.

expert testimony

- provide a special tool for capture of questions and answers
- brings steps together to unify experience
- bring in outside experience electronically
- pointers to potential problem areas
- prevention and detection of "excess creativity"
- analogy to experience: system learns from difference between its marking and a human's alternative marking
- analogy to tools: system is made of different specialized subsystems: NL understanding, planner, etc.
- analogy to information gathering: the working corpus – should we use unusual cases?
- analogy to manual and decision tree – should we expect non-expert to understand marking?
- analogy to structure of car – structure of question/answer
- accuracy and speed in both domains
- analogy to take car to body shop – submit run
- modularisation – engine was easy to remove
- for exam system, menu selecting is like the fault tree mentioned by Andy which ensures the efficiency for marking
- marking should follow certain schedules and algorithms, just like the timing of the car repair
- specific tools for particular exams, e.g. jargon

6. CONCLUSIONS AND FUTURE WORK

The participants in the workshop enjoyed working through the exercises. They were so engaged and enthusiastic that it was difficult to stop their work when it was time to go on to the next exercise. Future workshops should allow more time - at least 30 minutes - for the participants to brainstorm.

Thirteen participants produced 203 requirements in about two hours, including time for presentations and time for brainstorming. The expert testimony produced more requirements than any single creative trigger although many of them were of poor quality or not appropriate. For example, the meaning of "take car to body shop → submit run" is not clear. Taking the car to the body shop may be analogous to submitting a run for a computer program but the requirement for an automatic marker is unclear. Perhaps more time should be spent in future workshops to explain and give examples of creating requirements using analogies.

Not all of the requirements were comprehensible as written. Time should be scheduled for a verbal report by each group to ensure that the workshop leaders properly understand the written requirements.

The choice of the expert should be made carefully. The expert in this study tailored his presentation to encourage the participants to find analogies. A garrulous expert could consume a lot of time providing irrelevant or unhelpful detail.

Some participants thought that the order of the exercises should be reversed to stimulate maximum results while others felt that the order was correct. Nevertheless, all the groups came up with useful requirements. Future workshops could vary the order of the exercises to investigate a possible relationship between the order of the two exercises and quantity or quality of the generated requirements.

7. ACKNOWLEDGEMENTS

The authors thank James Robertson and Suzanne Robertson for the enjoyable and helpful tutorial they presented at RE'04.

The authors gratefully acknowledge the following members of the Open University READ group (Requirements Engineering and Design) for their participation in the Creative Requirements Workshop: Sarah Beecham, John Brier, Francis Chantree, Charles Haley, Jon Hall, Michael Jackson, Robin Laney, Zhi Li, Lun-Cheng Lin, Lucia Rapanotti, Janet van Der Linden, and Michel Wermelinger. Special thanks to Charles Haley, the OU's "Andy the auto mechanic."

The work reported in this study was partially supported by the European Community under the Innovation Society Technologies (IST) programme of the 6th Framework Programme for RTD - project ELeGI, contract IST-002205. This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of data appearing therein.

REFERENCES

[1] J. Robertson and S. Robertson, "Requirements and Creativity Tutorial," in *Requirements Engineering Conference RE 2004*. Kyoto: IEEE Computer Society, 2004.
[2] G. W. Furnas, S. Deerwester, S. T. Dumais, T. K. Landauer, R. A. Harshman, L. A. Streeter, and K. E. Lochbaum, "Information retrieval using a singular value decomposition model of latent semantic structure," *ACM*, pp. 465-480, 1988.
[3] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge," *Psychological Review*, vol. 104, pp. 211-240, 1997.
[4] T. K. Landauer, D. Laham, and P. W. Foltz, "Learning human-like knowledge by Singular Value Decomposition: A progress report," in *Advances in Neural Information Processing Systems 10*, M. I. Jordan, M. J. Kearns, and S. A. Solla, Eds. Cambridge: MIT Press, 1998, pp. 45-51.
[5] M. B. W. Wolfe, M. E. Schreiner, B. Rehder, D. Laham, P. W. Foltz, W. Kintsch, and T. K. Landauer, "Learning from text: Matching readers and texts by Latent Semantic Analysis," *Discourse Processes*, vol. 25, pp. 309-336, 1998.
[6] P. W. Foltz, D. Laham, and T. K. Landauer, "Automated Essay Scoring: Applications to Educational Technology," in *Proceedings of EdMedia '99*, 1999.
[7] C. Wu, M. Berry, S. Shivakumar, and J. McLarty, "Neural networks for full-scale protein sequence classification: Sequence encoding with singular value decomposition," *Machine Learning*, vol. 21, pp. 177-193, 1995.
[8] J. R. Skoyles, "Autistic language abnormality: Is it a second-order context learning defect? - The view from Latent Semantic Analysis," in *Proceedings Child Language Seminar*, I. Barriere, Chiat, S. G. Morgan, and B. Woll, Eds. London, 1999, pp. 1.
[9] R. S. Campbell and J. W. Pennebaker, "The secret life of pronouns: Flexibility in writing style and physical health," *Psychological Science*, vol. 14, pp. 60-65, 2003.
[10] D. J. Steinhart, "Summary Street: An intelligent tutoring system for improving student writing through the use of Latent Semantic Analysis," Department of Psychology, University of Colorado, Boulder, PhD Dissertation 2001.
[11] B. Lemaire and P. Dessus, "A system to assess the semantic content of student essays," *Journal of Educational Computing Research*, vol. 24, pp. 305-320, 2001.
[12] P. Wiemer-Hastings, K. Wiemer-Hastings, and A. C. Graesser, "Improving an intelligent tutor's comprehension of students with Latent Semantic Analysis," in *Artificial Intelligence in Education*, S. P. Lajoie and M. Vivet, Eds. Amsterdam: IOS Press, 1999.
[13] T. Miller, "Essay assessment with Latent Semantic Analysis," *Journal of Educational Computing Research*, vol. 28, 2003.
[14] D. Wade-Stein and E. Kintsch, "Summary Street: Interactive computer support for writing," University of Colorado, USA 2003.
[15] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, vol. 41, pp. 391-407, 1990.
[16] G. W. Furnas, L. M. Gomez, T. K. Landauer, and S. T. Dumais, "Statistical semantics: How can a computer use what people name things to guess what things people mean when they name things?," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: ACM, 1982, pp. 251-253.
[17] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, vol. 34, pp. 1-47, 2002.
[18] P. Wiemer-Hastings and A. C. Graesser, "Select-a-Kibitzer: A computer tool that gives meaningful feedback on student compositions," *Interactive Learning Environments*, vol. 8, pp. 149-169, 2000.
[19] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to Latent Semantic Analysis," *Discourse Processes*, vol. 25, pp. 259-284, 1998.