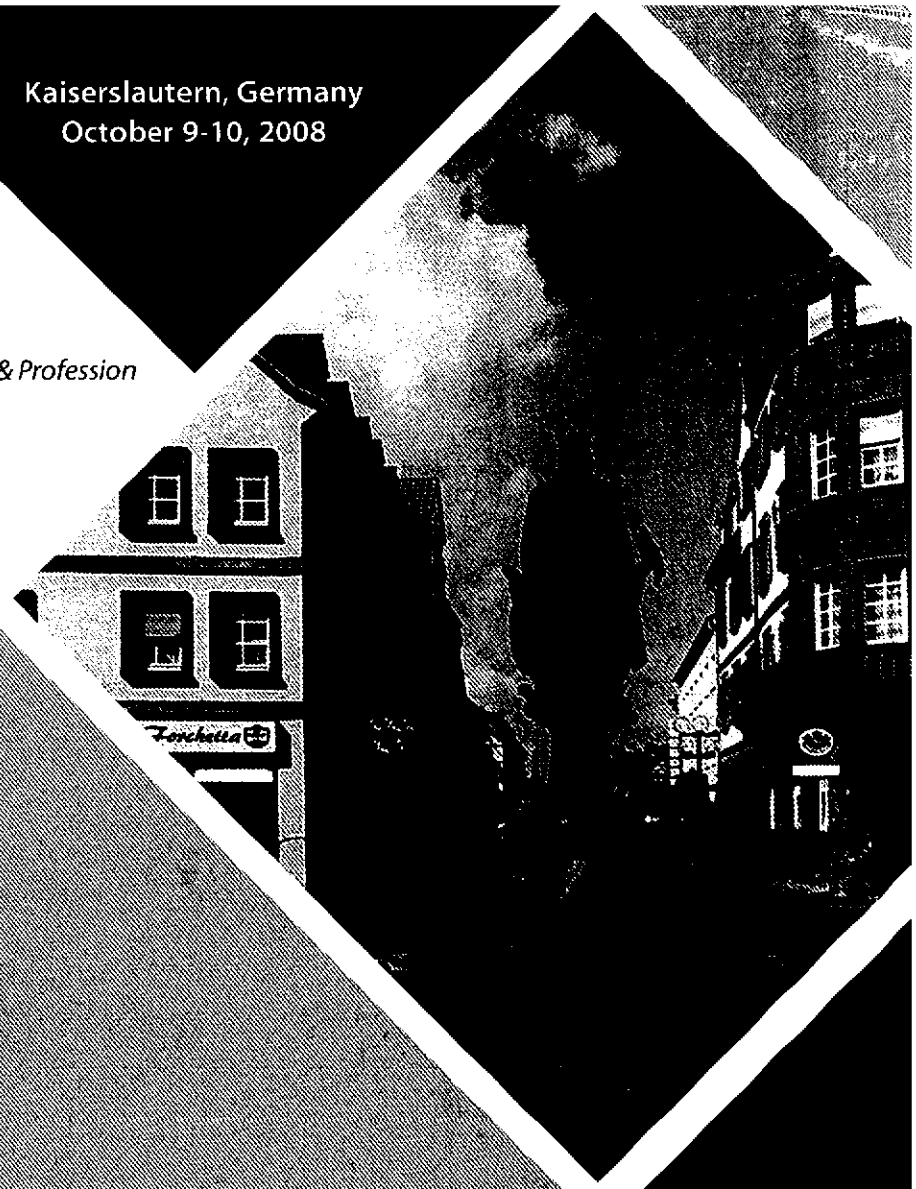
Kaiserslautern, Germany
October 9-10, 2008

Association for
Computing Machinery

Advancing Computing as a Science & Profession



ESEM'08

Proceedings of the 2008 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement

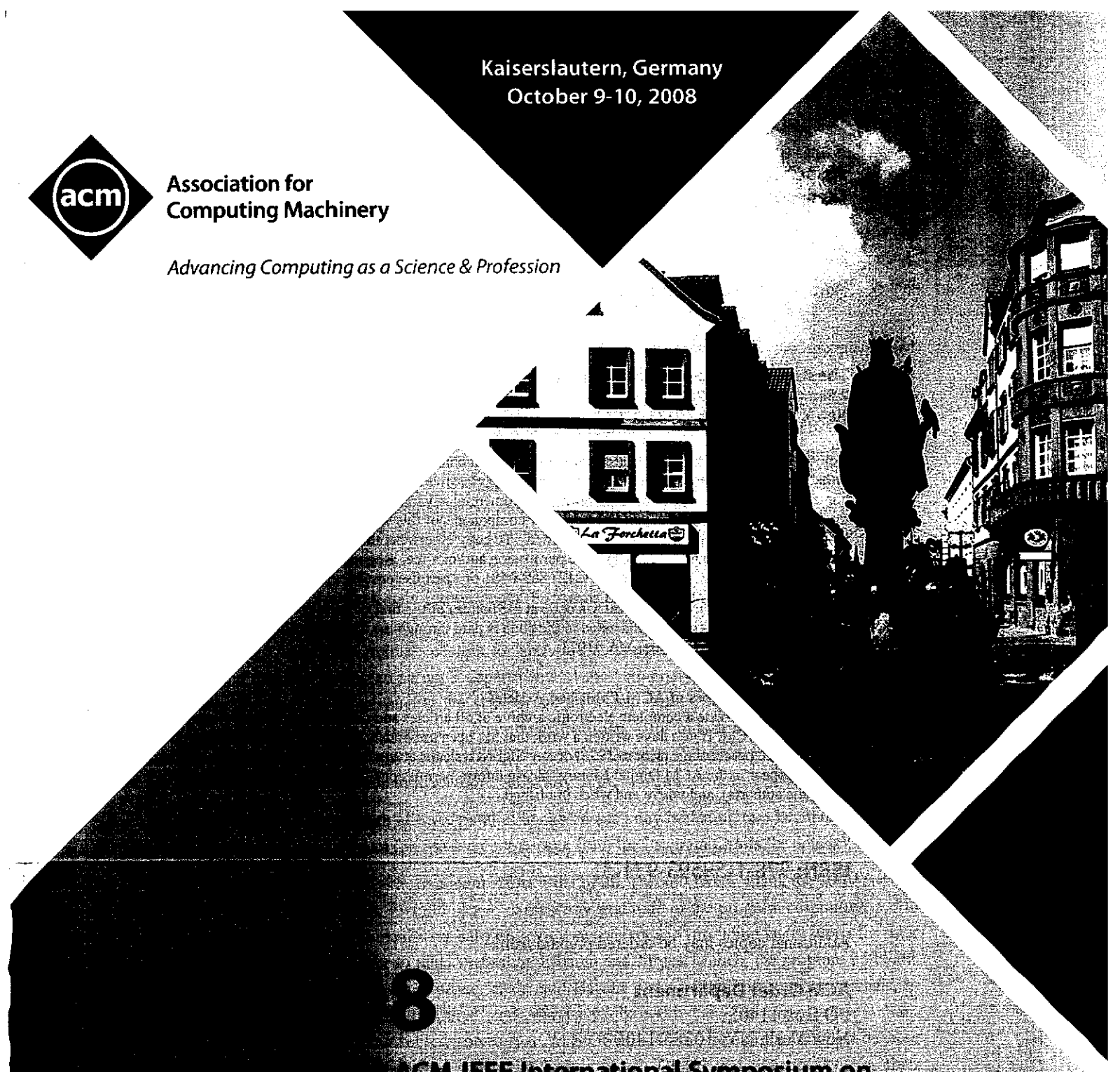
Sponsored by:

ACM SIGSOFT & IEEE CS

Supported by:

Siemens AG & Robert Bosch GmbH

Kaiserslautern, Germany
October 9-10, 2008



Association for
Computing Machinery

Advancing Computing as a Science & Profession

8

ACM-IEEE International Symposium on
re Engineering and Measurement

CS

t Bosch GmbH

Association for Computing Machinery

Advancing Computing as a Science & Profession

The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York, New York 10121-0701

ISBN: 978-1-59593-971-5

Additional copies may be ordered prepaid from:

ACM Order Department
PO Box 11405
New York, NY 10286-1405

Phone: 1-800-342-6626
(US and Canada)
+1-212-626-0500
(all other countries)
Fax: +1-212-944-1318
E-mail: acmhelp@acm.org

ACM Order Number 594087

Printed in the USA

Message from the General Chair

It is my great pleasure to welcome you to the Second International Symposium on Empirical Software Engineering and Measurement (ESEM 2008) in Kaiserslautern, Germany. Empirical software engineering and measurement are important sub-disciplines aimed at advancing software engineering towards an engineering discipline. They focus on measurement in order to better understand, control, and improve software development processes, and on empirical studies to better model products and processes and understand process products.

The 1st ESEM was held in Madrid in September 2007. Although this is only the second ESEM symposium, this series continues the long tradition of two high-quality conference series. ESEM resulted from the merger of the IEEE/ACM International Symposium on Empirical Software Engineering (ISESE) and the IEEE International Software Metrics Symposium (METRICS). ISESE was held five times, first in Japan in 2002 and last in Brazil in 2006. As its name suggests, ISESE's primary goal was to disseminate knowledge on and from experimental research on software engineering. METRICS is a long-standing conference series, which was held first in the United States in 1993 and last in Italy in 2005. METRICS' foremost objective was to disseminate knowledge on and from research on software metrics.

ESEM 2008 is part of the week-long Experimental Software Engineering International Week (ESEIW) in Kaiserslautern, Germany. It is an honor to welcome this conference to Kaiserslautern. Kaiserslautern has evolved into one of the premier centers for software & systems research in Europe. About 800 scientists and engineers – spread across 3 departments of the University of Kaiserslautern and several institutes – are working on advancing the state of the art in our field. The institutes include a Max-Planck Institute on Software Systems, two Fraunhofer Institutes – for Industrial Mathematics (ITWM) and Experimental Software Engineering (IESE) – and the German Research Institute for Artificial Intelligence (DFKI). I invite you to learn more about any of these institutes during your stay.

We have a full two-day program of three parallel tracks on each day including two keynotes, technical papers, short papers, and a panel (please see the message from the Program Chairs for details).

Putting together *ESEM 2008* has been a team effort. First of all, I would like to thank the authors of all submitted papers. Furthermore, I would like to express my gratitude to the program committee and to all external reviewers, who worked very hard on reviewing papers and providing suggestions for their improvements. I would also like to thank Sebastian Elbaum and Jürgen Münch, the program co-chairs Maria Teresa Baldassarre - the short papers chair, and Haruka Nakao - the poster program chair, as well as Andreas Jedlitschka, the treasurer, and Marcus Ciolkowski, the local arrangements chair. Finally, I would like to thank our society sponsor, ACM SIGSOFT, for their continued support of these successful meetings, and our industrial supporters Robert Bosch GmbH and Siemens AG for their financial support for the social events surrounding the conference.

I hope that you will find this program interesting and thought-provoking and that the symposium will provide you with a valuable opportunity to share ideas with other researchers and practitioners from institutions around the world.

Dieter Rombach
ESEM08 General Chair
University of Kaiserslautern & Fraunhofer IESE
Kaiserslautern, Germany

Message from the Program Co-Chairs

We would like to welcome you to Kaiserlautern, Germany, for the Second International Symposium on Empirical Software Engineering and Measurement – ESEM 2008. This year's technical program continues to advance the field of empirical software engineering and metrics with thoroughly reviewed technical papers complemented by two remarkable keynotes and a vibrant panel.

The call for research papers attracted over 100 submissions. To assess these submissions, the program committee members performed over 320 reviews and carried out dozens of online discussions. This process resulted in the selection of 28 papers that form our technical program. The papers cover a variety of topics, ranging from observational studies of programmers to measurement mechanisms for assessing technologies and organizations. In addition to the technical papers, short paper sessions and a poster exhibition will help to provide further coverage of ongoing work.

The keynote speakers and the panel will broaden and enrich the program by providing insights and reflections on our community strengths, weaknesses, and promising research directions. The keynotes will be given by Professor Mary Shaw from Carnegie Mellon University, USA, and Mr. Harald Hönninger, vice-president of Bosch corporate research division. The panel consists of highly renowned experts and promises interesting insights with respect to success stories, lessons learned from failures, and the next frontiers of empirical software engineering.

We would like to thank the authors who provided the content for the program, and to express our sincere gratitude to the program committee members and the external reviewers who volunteered time and resources to make this program possible. We would also like to thank the General Chair, Dieter Rombach, for his support and guidance, and Andreas Jedlitschka and Marcus Ciolkowski for the local organization of this conference and the maintenance of the website.

We hope that you will enjoy this program and that the symposium provides you with a valuable opportunity to share ideas with other researchers and practitioners.

Sebastian Elbaum
Program Co-Chair
University of Nebraska-Lincoln (USA)

Jürgen Münch
Program Co-Chair
Fraunhofer Institute for Experimental Software Engineering (Germany)

Table of Contents

Keynote Address

Session Chair: S. Elbaum *(University of Nebraska)*

Session 3A: From the Programmers' Trenches

Session Chair: H. Erdogmus *(NRC Institute for Information Technology)*

Session 3B: Inspections

Session Chair: D. Pfahl *(Simula Research Laboratory)*

Session 4A: Metrics and Methodology

Session Chair: M. Oivo *(University of Oulu)*

Session 4B: Faults and Failures

Session Chair: L. Williams *(North Carolina State University)*

Session 5A: Estimation Models II

Session Chair: M. Ciolkowski *(Fraunhofer Institute for Experimental Software Engineering)*

Session 5B: From the Manager's Trenches

Session Chair: P. Runeson *(Lund University)*

Short Papers

Session 1C: Evaluation and Comparison of Techniques and Models

Session Chair: M. Genero *(University of Castilla La Mancha)*

Session 2C: Empirical Studies of Processes and Products

Session Chair: D. Winkler *(Vienna University of Technology)*

Session 3C: Development of Predictive Models

Session Chair: O. Dieste *(Universidad Politécnica de Madrid)*

Session 4C: Experience in Process Improvement

Session Chair: T. Gorschek *(BTH)*

Session 5C: Empirical Evidence and Systematic Review

Session Chair: D. Caivano *(University of Bari)*

Posters

ESEM 2008 Organization

General Chair: Dieter Rombach - University of Kaiserslautern and Fraunhofer Institute for Experimental Software Engineering (Germany)

Program Chairs: Sebastian Elbaum - University of Nebraska-Lincoln (USA)
Jürgen Münch - Fraunhofer Institute for Experimental Software Engineering (Germany)

Steering Committee: Natalia Juristo - Universidad Politécnica de Madrid (Spain)
Sebastian Elbaum - University of Nebraska - Lincoln (USA)
James Miller - University of Alberta (Canada)
Jürgen Münch - Fraunhofer IESE (Germany)
Dieter Rombach - University of Kaiserslautern and Fraunhofer IESE (Germany)
Carolyn Seaman - University of Maryland at Baltimore County (USA)
Rick Selby - Northrop Grumman (USA)
Sira Vegas - Universidad Politécnica de Madrid (Spain)
Laurie Williams - North Carolina State University (USA)

Program Committee: Silvia Abrahao - Universidad Politecnica de Valencia (Spain)
Muhammad Ali Barbar - Lero, University of Limerick (Ireland)
Anneliese Amschler Andrews - University of Denver (USA)
James Andrews - University of Western Ontario (Canada)
Stefan Biffl - TU Vienna (Austria)
Lionel Briand - Simula Research Laboratory (Norway)
Luigi Buglione - Engineering.IT (Italy)
Gerardo Canfora - University of Sannio (Italy)
Jeff Carver - University of Alabama (USA)
Marcus Ciolkowski - University of Kaiserslautern (Germany)
Reidar Conradi - Norwegian University of Science and Technology (Norway)
Tore Dyba - Sintef (Norway)
Christof Ebert - Alcatel (France)
Khaled El Emam - University of Ottawa (Canada)
Hakan Erdogmus - NRC Institute for Information Technology (Canada)
Marcela Genero - University of Castilla La Mancha (Spain)
Tracy Hall - University of Hertfordshire (UK)
Lorin Hochstein - University of Nebraska-Lincoln (USA)
Martin Höst - Lund University (Sweden)
Frank Houdek - Daimlerchrysler (Germany)
Andreas Jedlitschka - Fraunhofer IESE (Germany)
Ross Jeffery - University of New South Wales (Australia)
Philip Johnson - University of Hawaii (USA)
Magne Jorgensen - Simula Research Laboratory (Norway)
Natalia Juristo - Universidad Politecnica de Madrid (Spain)
Shinji Kusumoto - Osaka University (Japan)

Program Committee (Continued):

Filippo Lanubile - University of Bari (Italy)
Jose Carlos Maldonado - Universidade de Sao Paulo (Brazil)
Kenichi Matsumoto - Nara Institute of Science and Technology (Japan)
Emilia Mendes - University of Auckland (New Zealand)
James Miller - University of Alberta (Canada)
Audris Mockus - Avaya Labs Research (USA)
Sandro Morasca - University degli Studi dell'Insubria (Italy)
Maurizio Morisio - Politecnico di Torino (Italy)
Nachi Nagappan - Microsoft (USA)
Markku Oivo - University of Oulu (Finland)
Dietmar Pfahl - Simula Research Laboratory & University of Oslo (Norway)
Mario Piattini - University of Castilla La Mancha (Spain)
Marc Roper - University of Strathclyde (UK)
Guenther Ruhe - University of Calgary (Canada)
Per Runeson - Lund University (Sweden)
Sreedevi Sampath - University of Maryland, Baltimore (USA)
Carolyn Seaman - University of Maryland, Baltimore (USA)
Rick Selby - Northrop Grumman (USA)
Martin Shepperd - Brunel University (UK)
Forrest Shull - Fraunhofer Center, Maryland (USA)
Janice Singer - National Research Council (Canada)
Harvey Siy - University of Nebraska (USA)
Dag Sjoberg - Simula Research Lababoratory (Norway)
Arie Van Deursen - Delft University of Technology (Nethrelands)
Sira Vegas - Universidad PolitÃ©cnica de Madrid (Spain)
June Verner - NICTA (Australia)
Laurie Williams - North Carolina State University (USA)
Claes Wohlin - Blekinge Institute of Technology (Sweden)

Finance Chair: Andreas Jedlitschka - Fraunhofer Institute for Experimental Software Engineering (Germany)

Local Arrangements Chair: Marcus Ciolkowski - Fraunhofer Institute for Experimental Software Engineering (Germany)

Publicity Chair: James Miller - University of Alberta (Canada)

Proceedings Chair: Marcus Ciolkowski - Fraunhofer Institute for Experimental Software Engineering (Germany)

Short Papers Chair: Teresa Baldassarre - Universita' degli Studiy di Bari (Italy)

Poster Chair: Haruka Nakao - Japan Manned Space Systems Cooperation (Japan)

Automatic Extraction of the Main Terminology used in Empirical Software Engineering through Text Mining Techniques

Francisco P. Romero, José A. Olivas
University of Castilla-La Mancha
Ciudad Real, Spain
{FranciscoP.Romero, JoseAngel.Olivas}@uclm.es

Marcela Genero, Mario Piattini
University of Castilla-La Mancha
Ciudad Real, Spain
{Marcela.Genero, Mario.Piattini}@uclm.es

ABSTRACT

The need for an explicit common terminology within Empirical Software Engineering (an ESE-Glossary of terms) was highlighted in the ISERN 2007 meeting [2]. The goal was to define a glossary of terms related to ESE based on an initial glossary published in http://lens-ese.cos.ufrj.br/wikiese. This initial glossary was built manually, based on expert knowledge. However, owing to the dynamic nature of the research works in ESE, this glossary must be dynamically updated with information extracted from the relevant documents in the research domain. Automation is, therefore, mandatory. We propose a text mining technique for the automatic extraction of the most relevant terms used in ESE documents. Our technique also provides the relationships between terms, with the degree of affinity between them. Our approach could, therefore, be useful in the improvement of the initial glossary of terms and in discovering relationships between terms.

Categories and Subject Descriptors

H.3.1 **[Information Storage and Retrieval]**: Content Analysis and Indexing – *Linguistic processing.*

D.2.0 **[Software Engineering]**: General. Standards.

General Terms

Documentation, Standardization.

1. INTRODUCTION

In this work the thirteenth volume of Empirical Software Engineering (EMSE) Journal [1] is used as a source of documents. In this preliminary work, it was decided to limit the number of reviewed documents to eleven, as this would be a sufficient quantity with which to carry out a preliminary study. These papers have been randomly selected. This is an ongoing work whose eventual purpose is to build an ESE Ontology.

2. FINDING RELEVANT TERMS IN ESE

We shall now describe the three main steps of which our text mining technique is composed. The result of this process is a hierarchical classification of terms.

1) Documents Preprocessing. It is first necessary to carry out a linguistic preprocess of the documents. The pre-processing scheme employs similar techniques, such as stemming and stop word removal, which are popularly used in the text mining community to extract terms for vector-space representation. The extraction process takes into account the presence of multiple-word terms. Abstract, references and acknowledgments are excluded since they contain redundant information. Finally, an index with these multi-word terms is built. This index contains the frequency of each term in each document processed.

2) Term Relevance. The automatic key terms extraction is performed through term ranking and by using term weighting measures. The frequency of the terms represents a good measure of the relevance in the domain. We propose the following extension of the classic tf-idf formula to calculate the relevance weight of a term i,:

$$w_i = \frac{docs(i) + \sum_{j \in D} \frac{tf(i,j) \cdot k}{|j|}}{|D|}$$

Thus, the relevance weight of a term *i* (w_i) is calculated by using the number of documents in which the term *i* appears (docs(i)), and its frequency in each document *j* (*tf(i,j)*). The normalization of the value is realized by using the number of processed documents (|D|). The formula also contains a multiplicity factor *k* (k = 1.000).

Table 1 shows the top ten individual and pairs of terms extracted from the afore mentioned EMSE Journal documents.

Table 1. Top ten of terms and pair of terms

Terms	Weight	Couple of terms	Weight
software	23.93	*software engineering*	9.51
study	16.93	empirical study	2.82
data	13.55	controlled experiment	1.82
engineering	13.18	*research interest*	1.74
project	12.40	*experiment design*	1.72
result	11.14	independent variable	1.63
experiment	10.27	*software development*	1.58
model	9.56	dependent variable	1.44
software eng.	9.51	research question	1.34
test	9.04	external validity	1.24

The terms in italics, which we consider to be relevant terms in the ESE field, do not appear in the published glossary. For that reason, our approach could, therefore, enrich the existent glossary.

ESEM'08, October 9–10, 2008, Kaiserslautern, Germany.
ACM 978-1-59593-971-5/08/10.

3) Term Relationships. This can be used to build a hierarchical classification of terms using the constructed index. An ontology, in this context, may be considered as a set of related trees in which each node represents a term.
In these trees, node A is a descendant of node B if the term A is "more concrete than" the term B, or if the term B is "more general than" the term A.
The fuzzy measure defined in [3] is used to obtain the generality degree (GD) between each pair of words contained in the document collection.

$$GD(A,B) = \frac{\text{Number of co - occurrences of A and B in the same subsection}}{\text{Number of occurences of B in the subsections}}$$

The T-DiCoR tool (Three Dimensional Conceptual Representation) could be used to visualize this hierarchical classifications of terms.

3. FUTURE WORK

In the near future, we shall expand the application of the text mining technique proposed in this paper to a greater volume of EMSE Journal documents. Publications in the ISESE and ESEM conferences should, moreover, be explored. Later, we shall obtain the relationships between terms and we go on build an ESE ontology.

4. REFERENCES

[1] Basili, V., Briand L. (eds). 2008 Empirical Software Engineering Journal. Vol 13. Springer.

[2] Travassos, G., Barker, M. 2007, Experimental Software Engineering Glossary of Terms, ISERN 2007.
Widyantoro D., Yen J. 2001 Incorporating fuzzy ontology of term relations in a search engine, Proceedings of the BISC Int. Workshop on Fuzzy Logic and the Internet, (Berkeley, USA), FLINT'01. 155-160.

Additional reviewers:

Martin Auer
Maria Teresa Baldassarre
Fabio Calefato
Marcio Delamaro
Oscar Dieste
Evgenia Egorova
Mohamed El-Attar
Fabiano Ferrari
Michael Gegick
Domenico Gendarmi
Stein Grimstad
Anna Grimán
Lauren Hayward
Sarah Smith Heckman
Jens Heidrich
Toan Huynh
Vigdis By Kampenes
Jacky Keung
Lucas Layman
Otávio Lemos
Jingyue Li
Teresa Mallardo
Rajwinder Kaur Panesar-Walawege
Francisco Pino
Daniel N. Port
Yonghee Shin
Ben Smith
Martín Soto
Tor Stålhane
Marco Torchiano
Adam Trendowicz
Irfan Ullah
Giuseppe Visaggio
Gursimran Walia
Byron Williams

Sponsors:

SIGSOFT

IEEE computer society

Supporters:

SIEMENS

Invented for life BOSCH

Author Index