Analysis of the Research Activities from ESSA Surveys 2009 COPSRO Project

(grant PTDC/EIA/69988/2006)

Susana Nascimento[1]
[1]Computer Science Department and Center for Artificial Intelligence (CENTRIA)
Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa
Caparica, Portugal

February 10, 2010

Contents

List of Tables

List of Figures

1 Introduction

This report presents results of the analysis of three research organizations in the framework of the project *Computational Ontology Profiling of Scientific Research Organization* (COPSRO), the main goal of which is to develop a method to represent a Computer Science organization such as a university department over the ACM Computing Classification System (ACM-CCS) (http:// www.acm.org/class/1998/ccs98.html). Our approach involves the following steps:

1. Surveying the members of the organization regarding the ACM-CCS topics they work on in order to derive their research profiles; this may be supplemented with indication of the degree of success achieved (publication in a good journal, grant award, etc.);
2. Deriving thematic similarity between ACM-CCS topics according to the research profiles resulting from the survey and then clustering them;
3. Mapping the fuzzy clusters to the ACM-CCS taxonomy and lifting them to higher ranks in a parsimonious way by minimizing the penalty for emerging head subjects, gaps and offshoots;
4. Aggregating the results from different clusters and, subsequently, different organizations over the taxonomy;
5. Interpreting the results.

An electronic survey tool has been developed, ESSA (available at https://copsro.di.fct.unl.pt/), which in its baseline version, a respondent is asked to select up to six topics among the third layer nodes of the ACM-CCS tree[1] and assign each with a percentage expressing the proportion of the topic in the total of the respondent's research activity for the last five years (i.e. current year + the previous four years).

The set of profiles supplied by respondents forms an $N \times M$ matrix F where N is the number of ACM-CCS topics involved in the profiles and M the number of respondents. Each column of F is a fuzzy membership function.

During the year of 2009, three surveys have been conducted over the ESSA e-surveying tool in the following university institutions:

- Centre for Artificial Intelligence, Universidade Nova de Lisboa, Portugal (CENTRIA-UNL);
- Department of Computer Science and Information Systems, Birkbeck, University of London, London, UK (CSIS-BKUL);
- Department of Computer Science, Instituto Superior de Engenharia do Porto, Portugal (DEI-ISEP).

Table 1 summarizes the data samples obtained in each of the surveys

The collected data has been analysed according to the following 2-phase generalization procedure:

[1] The total number of 3rd layer ACM-CCS topics is equal to 318.

Table 1: Sizes of ESSA'09 surveys' data

	N. of Contacted Respondents	N. of Participating Respondents	N. of 3rd Layer ACM-CCS Topics Covered
CENTRIA-UNL	23	16	46/318
CSIS-BKUL	40	19	54/318
DEI-ISEP	30	23	65/318

1. Build thematic similarities and fuzzy clustering by the one-by-one fuzzy thematic cluster extraction spectral algorithm, referred to as ADDItive Fuzzy Spectral clustering (ADDI-FS) method [1].

2. Mapping and visualizing clusters to the ACM-CCS taxonomy and lifting them to higher ranks in a parsimonious way by minimizing the penalty for emerging "head subjects", "gaps" and "offshoots", (Parsimonious Lifting method) [2].

The following aspects should be pointed out about the 2-phase generalization Cluster-Lift procedure:

- The ADDI-FS method is a version of approximate clustering modified with the spectral clustering approach to make use of the Laplacian data transformation which has proven an effective tool to sharpen the cluster structure hidden in the data. The method involves a number of model-based stopping conditions. For a detailed description of the ADDI–FS method, including its application to affinity and community structure data along with experimental comparisons with other fuzzy clustering methods, consult [1].

- To generalize the contents of a thematic cluster, we lift it to higher ranks of the taxonomy so that if all or almost all children of a node in an upper layer belong to the cluster, then the node itself is taken to represent the cluster at this higher level of the ACM-CCS taxonomy. Such lifting can be done differently, leading to different portrayals of the cluster on ACM-CCS tree depending on the relative weights of the events taken into account.

- A major event is the so-called "head subject", a taxonomy node covering (some of) leaves belonging to the cluster, so that the cluster is represented by a set of "head subjects". The penalty of the representation to be minimized is proportional to the number of "head subjects" so that the smaller that number the better. Yet the "head subjects" cannot be lifted too high in the tree because of the penalties for associated events, the cluster"gaps" and "offshoots". The "gaps" are head subject's children topics that are not included in the cluster. An offshoot is a taxonomy leaf node that is a head subject not lifted.

- The total count of "head subjects", "gaps" and "offshoots", each weighted by both the penalties and leaf memberships, is used for scoring the extent of the cluster misfit needed for lifting a grouping of research topics over the classification tree. The smaller the score, the more parsimonious the lift and the better the fit. Depending on

the relative weighting of “gaps”, “offshoots” and multiple “head subjects”, different lifts can minimize the total misfit.

- Altogether, the set of topic clusters together with their optimal “head subjects”, “offshoots” and “gaps” constitute a *parsimonious representation of the organization*.

In the following sections we present our results for each of the three organizations: CENTRIA (Section 2), CSIS-BKUL (Section 3), and DEI-ISEP (Section 4).

2 Analysis of the Research Center CENTRIA-UNL

2.1 Description of the Clusters

Resulting of the application of the ADDI-FS algorithm only two clusters have been extracted, after which the residual similarity matrix turned negative definite to halt the process.

Table 2 presents the two fuzzy clusters obtained from ADDI-FS algorithm on the analysis of CENTRIA-UNL data set. The clusters' membership values are sorted in the descending order. CENTRIA's two clusters contributions total is about 50%.

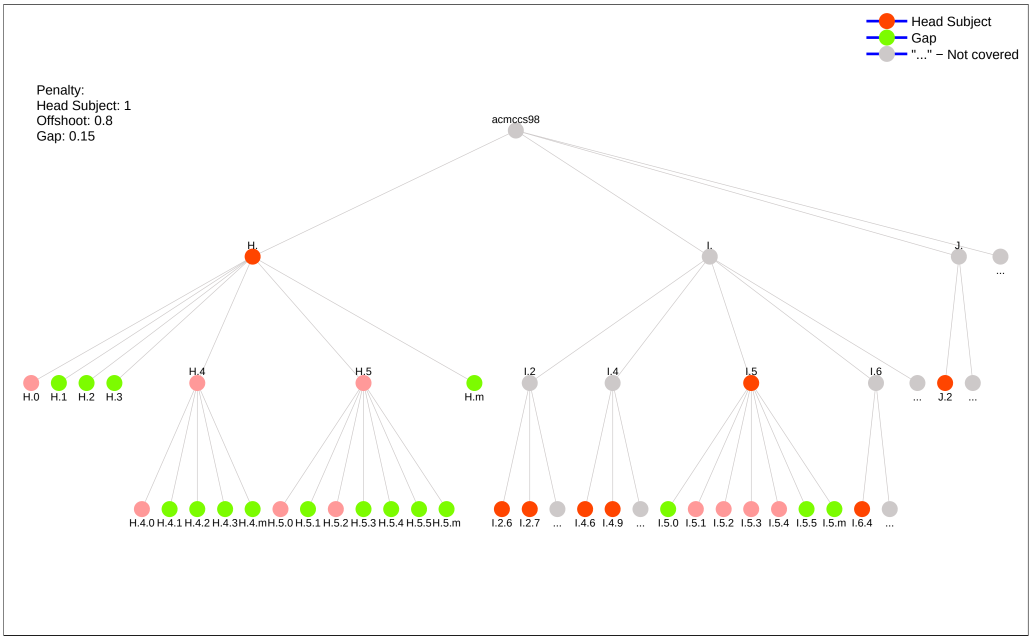
2.2 Description of the Mapping Results

Results of the cluster lifting procedure with penalties for head subjects (h), offshoots (o) and gaps g) of: $h = 1$, $o = 0.8$, and $g = 0.15$, are illustrated in Figure 1 (cluster 1) and Figure 2 (cluster 2).

Tables 3 and 4 summarize the parsimonious representation of CENTRIA in terms of head subjects, offshoots, and gaps.

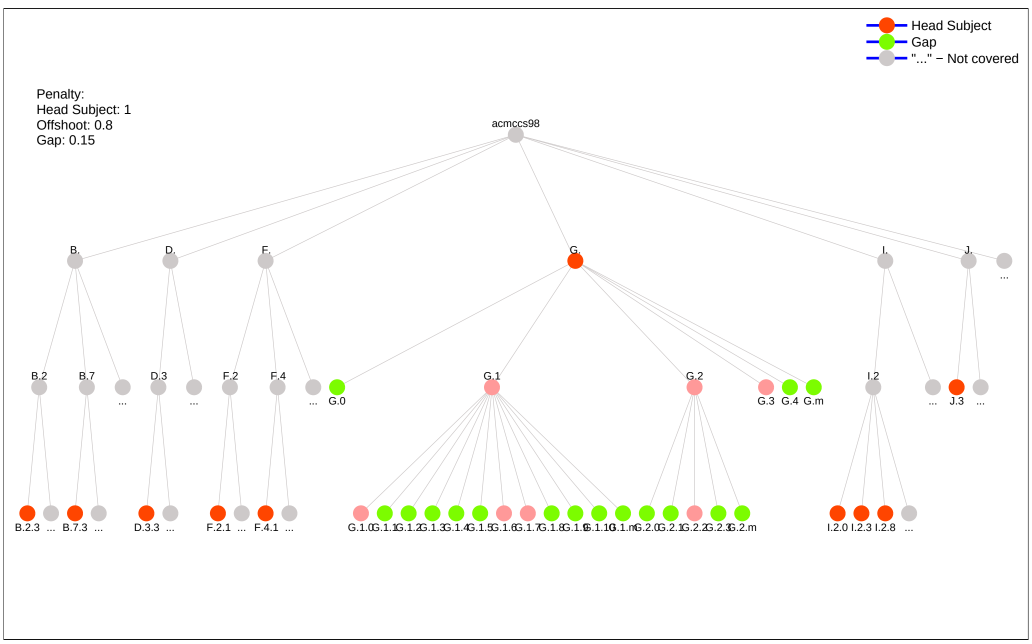
Table 2: ADDI-FS results at data of similarity of research topics in CENTRIA-UNL

Cluster 1		
Eigenvalue	46.50	
Contribution	35.2%	
Intensity	5.57	
Weight	31.04	
Membership	Code	Topic
0.69911	I.5.3	Clustering
0.3512	I.5.4	Applications in I.5 PATTERN RECOGNITION
0.27438	J.2	PHYSICAL SCIENCES AND ENGINEERING (Applications in)
0.1992	I.4.9	Applications in I.4 IMAGE PROCESSING AND COMPUTER VISION
0.1992	I.4.6	Segmentation
0.19721	I.2.6	Learning
0.17478	H.5.2	User Interfaces
0.17478	I.6.4	Model Validation and Analysis in I.6 SIMULATION AND MODELING
0.16689	I.2.7	Natural Language Processing
0.16689	I.5.1	Models in I.5 PATTERN RECOGNITION
0.14453	I.5.2	Design Methodology (Classifiers)
0.13646	H.5.0	General in H.5 INFORMATION INTERFACES AND PRESENTATION
0.13646	H.0	GENERAL in H. Information Systems
0.13646	H.4.0	General in H.4 INFORMATION SYSTEMS APPLICATIONS
0.02867	I.2.11	Distributed Artificial Intelligence
Cluster 2		
Contribution	15.2%	
Eigenvalue	32.90	
Intensity	4.52	
Weight	20.41	
Membership	Code	Topic
0.46756	J.3	LIFE AND MEDICAL SCIENCES (Applications in)
0.40619	I.2.8	Problem Solving, Control Methods, and Search
0.34435	F.2.1	Numerical Algorithms and Problems
0.32681	F.4.1	Mathematical Logic
0.30067	G.1.6	Optimization
0.25967	D.3.3	Language Constructs and Features
0.23748	G.2.2	Graph Theory
0.18722	G.3	PROBABILITY AND STATISTICS
0.17359	B.2.3	Reliability, Testing, and Fault-Tolerance
0.17359	B.7.3	Reliability and Testing
0.17203	I.2.0	General in I.2 ARTIFICIAL INTELLIGENCE
0.1537	G.1.0	General in G.1 NUMERICAL ANALYSIS
0.11827	I.2.3	Deduction and Theorem Proving
0.10195	G.1.7	Ordinary Differential Equations
0.06175	K.2	HISTORY OF COMPUTING
0.00726	D.1.6	Logic Programming



Mapping of CENTRIA Cluster 1 on ACM−CCS taxonomy

Figure 1: Mapping of CENTRIA cluster 1 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.



Mapping of CENTRIA Cluster 2 on ACM−CCS taxonomy

Figure 2: Mapping of CENTRIA cluster 2 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.

Table 3: Parsimonious Representation of CENTRIA Cluster 1

	HEAD SUBJECTS
H.	Information Systems
I.5	PATTERN RECOGNITION
	OFFSHOTS
I.2.6	Learning
I.2.7	Natural Language Processing
I.4.6	Segmentation
I.4.9	Applications
I.6.4	Model Validation and Analysis
J.2	PHYSICAL SCIENCES AND ENGINEERING
	GAPS
H.1	MODELS AND PRINCIPLES
H.2	DATABASE MANAGEMENT
H.3	INFORMATION STORAGE AND RETRIEVAL
H.4.1	Office Automation
H.4.2	Types of Systems
H.4.3	Communications Applications
H.4.m	Miscellaneous in H.4 - INFORMATION SYSTEMS APPLICATIONS
H.5.1	Multimedia Information Systems
H.5.3	Group and Organization Interfaces
H.5.4	Hypertext/Hypermedia
H.5.5	Sound and Music Computing
H.5.m	Miscellaneous in H.5 - INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)
H.m	MISCELLANEOUS in H. - Information Systems
I.5.0	General in I.5 - PATTERN RECOGNITION
I.5.5	Implementation
I.5.m	Miscellaneous in I.5 - PATTERN RECOGNITION

Table 4: Parsimonious Representation of CENTRIA Cluster 2

	HEAD SUBJECTS
G.	Mathematics of Computing
	OFFSHOOTS
B.2.3	Reliability, Testing, and Fault-Tolerance
B.7.3	Reliability and Testing
D.3.3	Language Constructs and Features
F.2.1	Numerical Algorithms and Problems
F.4.1	Mathematical Logic
I.2.0	General in I.2 - ARTIFICIAL INTELLIGENCE
I.2.3	Deduction and Theorem Proving
I.2.8	Problem Solving, Control Methods, and Search
J.3	LIFE AND MEDICAL SCIENCES
	GAPS
G.0	GENERAL in G. - Mathematics of Computing
G.1.1	Interpolation
G.1.2	Approximation
G.1.3	Numerical Linear Algebra
G.1.4	Quadrature and Numerical Differentiation
G.1.5	Roots of Nonlinear Equations
G.1.8	Partial Differential Equations
G.1.9	Integral Equations
G.1.10	Applications
G.1.m	Miscellaneous in G.1 - NUMERICAL ANALYSIS
G.2.0	General in G.2 - DISCRETE MATHEMATICS
G.2.1	Combinatorics
G.2.3	Applications
G.2.m	Miscellaneous in G.2 - DISCRETE MATHEMATICS
G.4	MATHEMATICAL SOFTWARE
G.m	MISCELLANEOUS in G. - Mathematics of Computing

Table 5: ADDI-FS results at data of similarity of research topics in university department CSIS-BKUL

Cluster 1		
Eigenvalue	37.44	
Contribution	26.7%	
Intensity	5.26	
Weight	27.68	
Membership	Code	Topic
0.43055	K.2	HISTORY OF COMPUTING
0.39255	D.2.11	Software Architectures
0.35207	C.2.4	Distributed Systems
0.3412	I.2.11	Distributed Artificial Intelligence
0.3335	K.7.3	Testing, Certification, and Licensing
0.30491	D.2.1	Requirements/Specifications in D.2 Software Engineering
0.27437	D.2.2	Design Tools and Techniques in D.2 Software Engineering
0.24126	C.3	SPECIAL-PURPOSE AND APPLICATION-BASED SYSTEMS
0.19525	D.1.6	Logic Programming
0.19525	D.2.7	Distribution, Maintenance, and Enhancement in D.2 Software Engineering
Cluster 2		
Contribution	13.4%	
Eigenvalue	26.65	
Intensity	4.43	
Weight	19.60	
Membership	Code	Topic
0.66114	J.1	ADMINISTRATIVE DATA PROCESSING
0.29567	K.6.1	Project and People Management in K.6
0.29567	K.6.0	General in K.6 MANAGEMENT OF COMPUTING AND INF. SYSTEMS
0.29567	H.4.m	Miscellaneous in H.4 INF. SYSTEMS APPLICATIONS
0.29567	J.7	COMPUTERS IN OTHER SYSTEMS
0.2696	J.4	SOCIAL AND BEHAVIORAL SCIENCES
0.16271	J.3	LIFE AND MEDICAL SCIENCES
0.14985	G.2.2	Graph Theory
0.14593	I.5.3	Clustering
0.12307	I.6.4	Model Validation and Analysis
0.10485	I.6.5	Model Development

3 Analysis of the University Department CSIS-BKUL

3.1 Description of the Clusters

Tables 5 and 6 present the four fuzzy clusters obtained from ADDI-FS algorithm on the analysis of CSIS-BKUL data set. Cluster membership values are sorted in the descending order. The clusters' contributions total to about 60%.

3.2 Description of the Mapping Results

Results of the cluster lifting procedure with penalties for "'head subjects"' (h), offshoots (o) and gaps g) of: $h = 1$, $o = 0.8$, and $g = 0.15$, are illustrated in Figures 3-6 for clusters

Table 6: ADDI-FS results at data of similarity of research topics in university department CSIS-BKUL

Cluster 3		
Contribution	18.9%	
Eigenvalue	24.31	
Intensity	4.83	
Weight	23.31	
Membership	Code	Topic
0.613	E.2	DATA STORAGE REPRESENTATIONS
0.55728	I.0	GENERAL in I. Computing Methodologies
0.55728	H.0	GENERAL in H. Information Systems
Cluster 4		
Contribution	3.7%	
Eigenvalue	19.05	
Intensity	3.20	
Weight	10.26	
Membership	Code	Topic
0.35713	I.2.4	Knowledge Representation Formalisms and Methods
0.35636	F.4.1	Mathematical Logic
0.29495	F.2.0	General in F.2 ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY
0.28713	I.5.0	General in I.5 PATTERN RECOGNITION
0.28169	I.2.6	Learning
0.25649	K.3.1	Computer Uses in Education
0.24848	I.4.0	General in I.4 IMAGE PROCESSING AND COMPUTER VISION
0.24083	F.4.0	General in F.4 MATHEMATICAL LOGIC AND FORMAL LANGUAGES
0.18644	H.2.8	Database Applications
0.17707	H.2.1	Logical Design
0.17029	I.2.3	Deduction and Theorem Proving
0.15727	E.1	DATA STRUCTURES
0.15306	I.5.3	Clustering
0.14976	F.2.2	Nonnumerical Algorithms and Problems
0.14809	I.2.8	Problem Solving, Control Methods, and Search
0.14809	I.2.0	General in I.2 ARTIFICIAL INTELLIGENCE

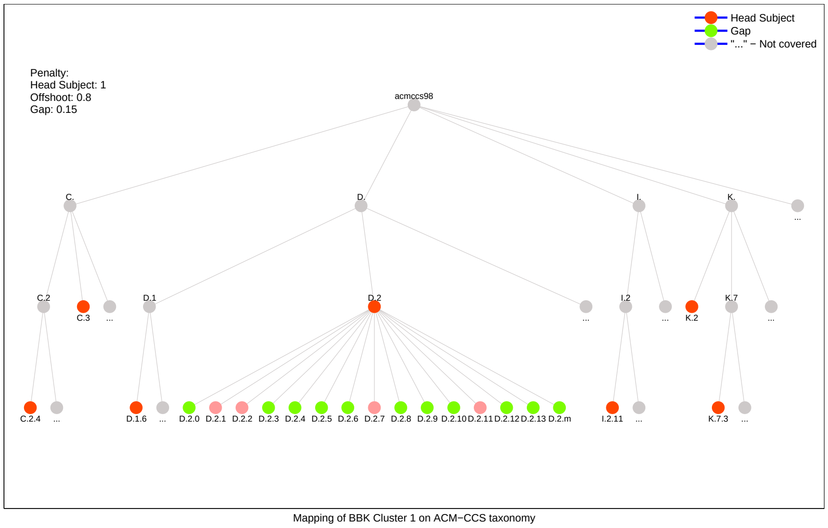


Figure 3: Mapping of CSIS-BBK cluster 1 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.

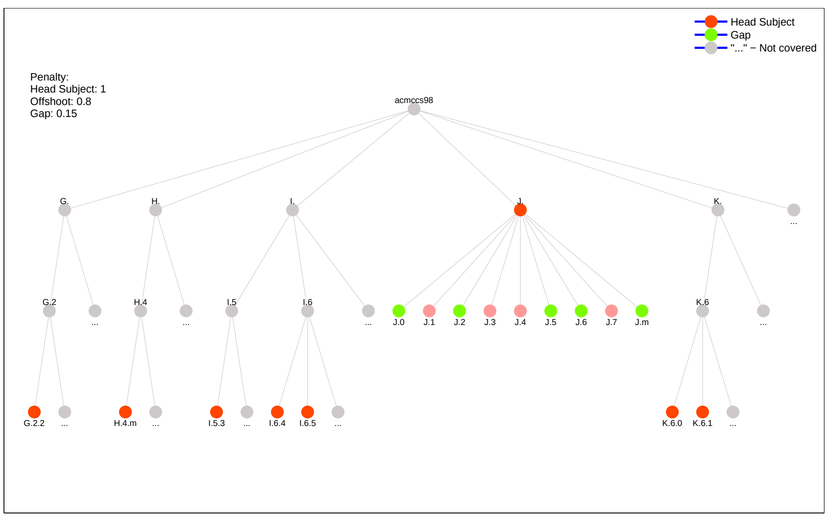


Figure 4: Mapping of CSIS-BBK cluster 2 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.

Cluster1 to Cluster4, respectively.

Tables 7-10 summarize the parsimonious representation of the CSIS-BKUL department in terms of head subjects, offshoots and gaps.

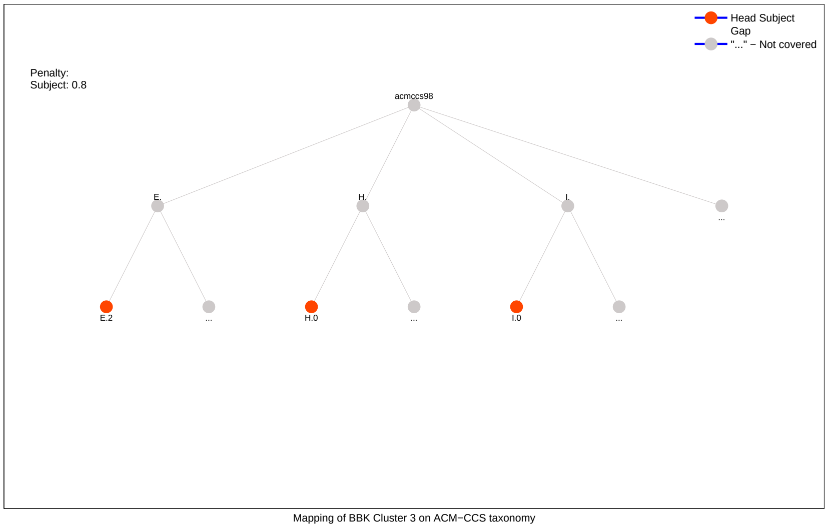


Figure 5: Mapping of CSIS-BBK cluster 3 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.

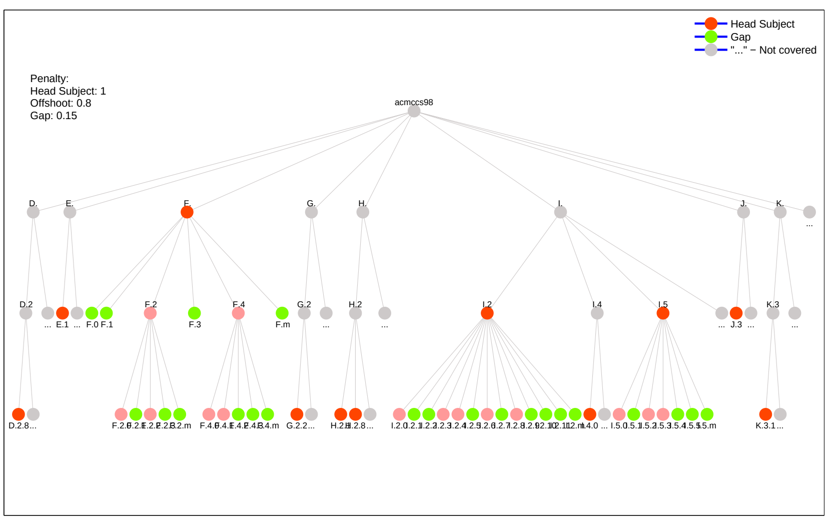


Figure 6: Mapping of CSIS-BBK cluster 4 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.

Table 7: Parsimonious Representation of CSIS-BKUL Cluster 1

	HEAD SUBJECTS
D.2	SOFTWARE ENGINEERING
	OFFSHOTS
C.2.4	Distributed Systems
C.3	SPECIAL-PURPOSE AND APPLICATION-BASED SYSTEMS
D.1.6	Logic Programming
I.2.11	Distributed Artificial Intelligence
K.2	HISTORY OF COMPUTING
K.7.3	Testing, Certification, and Licensing
	GAPS
D.2.0	General in D.2 - SOFTWARE ENGINEERING
D.2.3	Coding Tools and Techniques
D.2.4	Software/Program Verification
D.2.5	Testing and Debugging
D.2.6	Programming Environments
D.2.8	Metrics
D.2.9	Management
D.2.10	Design
D.2.12	Interoperability
D.2.13	Reusable Software
D.2.m	Miscellaneous in D.2 - SOFTWARE ENGINEERING

Table 8: Parsimonious Representation of CSIS-BKUL Cluster 2

	HEAD SUBJECTS
J.	Computer Applications
	OFFSHOTS
G.2.2	Graph Theory
H.4.m	Miscellaneous in H.4 - INFORMATION SYSTEMS APPLICATIONS
I.5.3	Clustering
I.6.4	Model Validation and Analysis
I.6.5	Model Development
K.6.0	General in K.6 - MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS
K.6.1	Project and People Management
	GAPS
J.0	GENERAL in J. - Computer Applications
J.2	PHYSICAL SCIENCES AND ENGINEERING
J.5	ARTS AND HUMANITIES
J.6	COMPUTER-AIDED ENGINEERING
J.m	MISCELLANEOUS in J. - Computer Applications

Table 9: Parsimonious Representation of CSIS-BKUL Cluster 3

	SUBJECTS
E.2	DATA STORAGE REPRESENTATIONS
H.0	GENERAL in H. - Information Systems
I.0 -	GENERAL in I. - Computing Methodologies

Table 10: Parsimonious Representation of CSIS-BKUL Cluster 4

	HEAD SUBJECTS
F.	Theory of Computation
I.2	ARTIFICIAL INTELLIGENCE
I.5	PATTERN RECOGNITION
	OFFSHOTS
D.2.8	Metrics
E.1	DATA STRUCTURES
G.2.2	Graph Theory
H.2.1	Logical Design
H.2.8	Database Applications
I.4.0	General in I.4 - IMAGE PROCESSING AND COMPUTER VISION
J.3	LIFE AND MEDICAL SCIENCES
K.3.1	Computer Uses in Education
	GAPS
F.0	GENERAL in F. - Theory of Computation
F.1	COMPUTATION BY ABSTRACT DEVICES
F.2.1	Numerical Algorithms and Problems
F.2.3	Tradeoffs between Complexity Measures
F.2.m	Miscellaneous in F.2 - ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY
F.3	LOGICS AND MEANINGS OF PROGRAMS
F.4.2	Grammars and Other Rewriting Systems
F.4.3	Formal Languages
F.4.m	Miscellaneous in F.4 - MATHEMATICAL LOGIC AND FORMAL LANGUAGES
F.m	MISCELLANEOUS in F. - Theory of Computation
I.2.1	Applications and Expert Systems
I.2.2	Automatic Programming
I.2.5	Programming Languages and Software
I.2.7	Natural Language Processing
I.2.9	Robotics
I.2.10	Vision and Scene Understanding
I.2.11	Distributed Artificial Intelligence
I.2.m	Miscellaneous in I.2 - ARTIFICIAL INTELLIGENCE
I.5.1	Models
I.5.4	Applications
I.5.5	Implementation
I.5.m	Miscellaneous in I.5 - PATTERN RECOGNITION

Table 11: ADDI-FS results at data of similarity of research topics in university department DEI-ISEP.

Cluster 1		
Eigenvalue	26.01	
Contribution	30.19%	
Intensity	4.57	
Weight	20.88	
Membership	Code	Topic
0.563	D.2.9	Management
0.411	D.2.1	Requirements/Specifications
0.323	D.2.12	Interoperability
0.323	J.3	LIFE AND MEDICAL SCIENCES
0.276	D.2.2	Design Tools and Techniques
0.249	D.2.10	Design
0.236	D.2.11	Software Architectures
0.228	K.7.3	Testing, Certification, and Licensing
0.174	D.2.m	Miscellaneous
0.144	D.2.13	Reusable Software
0.091	D.1.5	Object-oriented Programming
0.019	I.6.8	Types of Simulation
0.019	E.1	DATA STRUCTURES
Cluster 2		
Contribution	22.36%	
Eigenvalue	20.60	
Intensity	4.24	
Weight	17.97	
Membership	Code	Topic
0.479	C.2.4	Distributed Systems
0.479	H.3.4	Systems and Software
0.333	D.3.3	Language Constructs and Features
0.333	F.3.2	Semantics of Programming Languages
0.276	D.4.1	Process Management
0.276	F.3.3	Studies of Program Constructs
0.236	D.3.1	Formal Definitions and Theory
0.236	D.3.4	Processors
0.203	D.1.6	Logic Programming
0.118	A.1	INTRODUCTORY AND SURVEY

4 Analysis of the University Department DEI-ISEP

4.1 Description of the Clusters

Tables 11 and 12 present the four fuzzy clusters obtained from ADDI-FS algorithm on the analysis of DEI-ISEP data set. Cluster membership values are sorted in the descending order.

Table 12: ADDI-FS results at data of similarity of research topics in university department DEI-ISEP.

Cluster 3		
Eigenvalue	14.03	
Contribution	9.65%	
Intensity	3.44	
Weight	11.80	
Membership	Code	Topic
0.628	I.5.3	Clustering
0.339	H.2.8	Database Applications
0.305	H.3.3	Information Search and Retrieval
0.294	H.2.1	Logical Design
0.277	I.5.5	Implementation
0.277	I.5.4	Applications
0.249	I.2.5	Heterogeneous Databases
0.226	J.2	PHYSICAL SCIENCES AND ENGINEERING
0.138	I.2.4	Knowledge Representation Formalisms and Methods
0.125	H.3.1	Content Analysis and Indexing
0.05	I.6.5	Model Development
0.038	F.3.1	Specifying and Verifying and Reasoning about Programs
0.036	I.2.6	Learning
0.033	I.2.1	Applications and Expert Systems
0.031	I.6.7	Simulation Support Systems
0.027	I.6.1	Simulation Theory
0.027	J.6	COMPUTER-AIDED ENGINEERING
0.027	D.2.4	Software/Program Verification
0.014	K.4.2	Social Issues
Cluster 4		
Contribution	8.34%	
Eigenvalue	11.77	
Intensity	3.31	
Weight	10.97	
Membership	Code	Topic
0.89	E.m	MISCELLANEOUS in DATA
0.38	I.2.m	Miscellaneous in ARTIFICIAL INTELLIGENCE
0.133	I.2.5	Programming Languages and Software
0.058	I.2.6	Learning
0.056	I.2.1	Applications and Expert Systems
0.051	I.1.2	Algorithms
0.036	H.1.2	User/Machine Systems
0.036	J.7	COMPUTERS IN OTHER SYSTEMS
0.036	D.1.6	Logic Programming
0.03	I.2.4	Knowledge Representation Formalisms and Methods
0.03	J.6	COMPUTER-AIDED ENGINEERING
0.023	I.6.5	Model Development
0.02	J.4	SOCIAL AND BEHAVIORAL SCIENCES
0.017	I.2.11	Distributed Artificial Intelligence
0.017	F.3.1	Specifying and Verifying and Reasoning about Programs
0.016	F.4.3	Formal Languages
0.016	I.6.7	Simulation Support Systems
0.012	D.2.4	Software/Program Verification
0.011	I.6.1	Simulation Theory

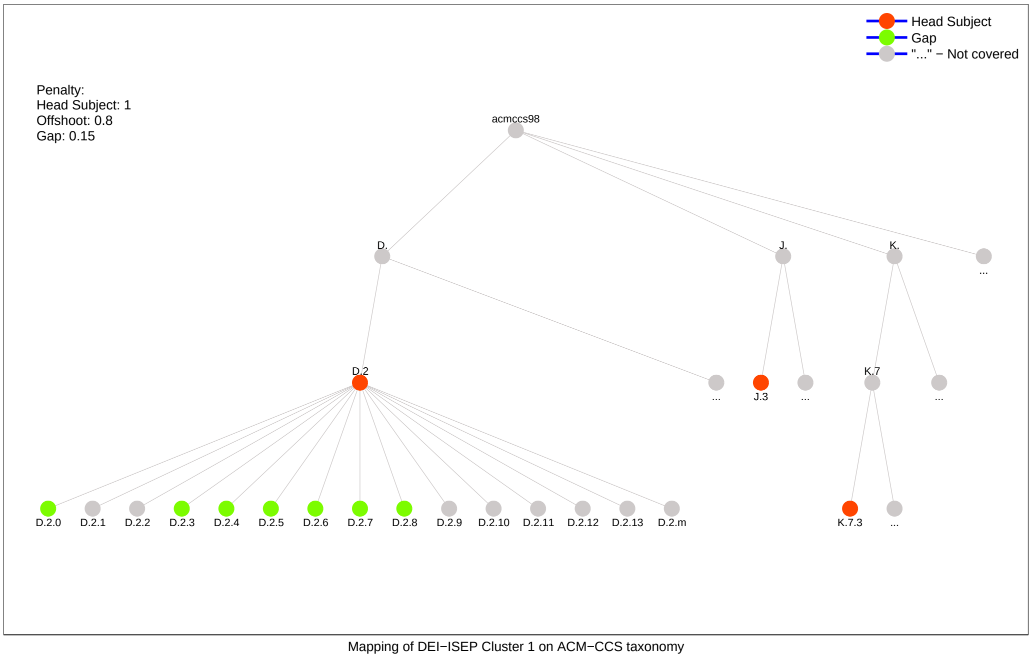
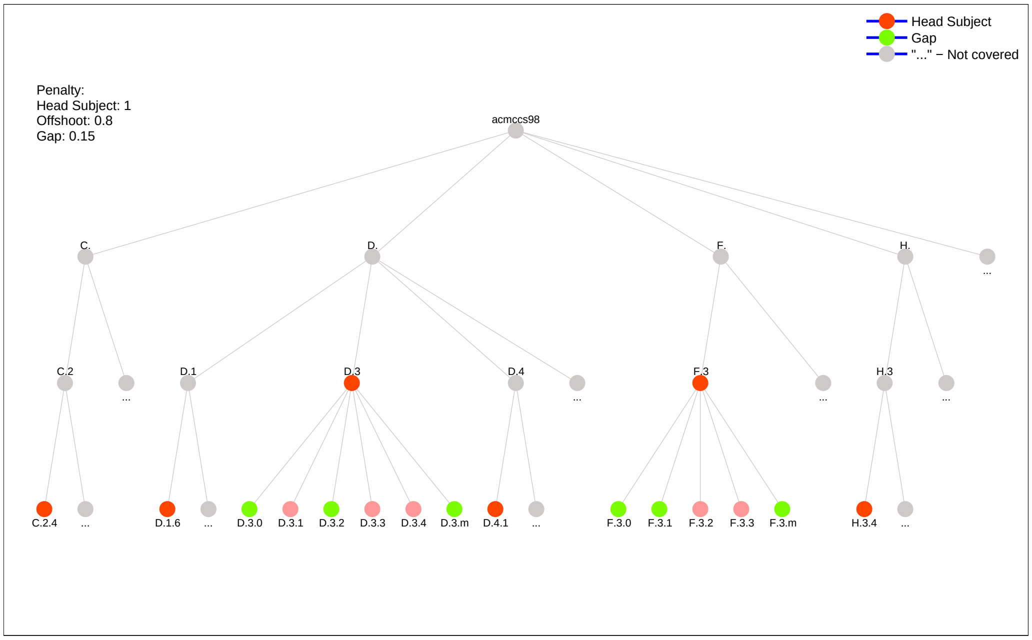


Figure 7: Mapping of DEI-ISEP cluster 1 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.

4.2 Description of the Mapping Results

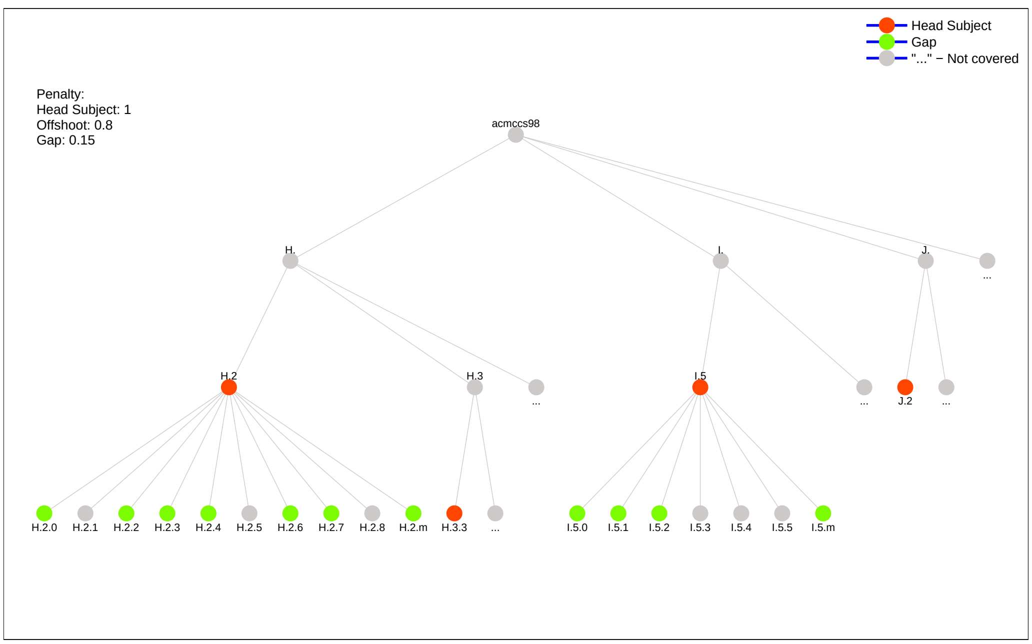
Results of the cluster lifting procedure with penalties for head subjects (h), offshoots (o) and gaps g) of: $h = 1$, $o = 0.8$, and $g = 0.15$, are illustrated in Figures 7-10 for clusters Cluster1 to Cluster4, respectively.

Tables 13-16 summarize the parsimonious representation of DEI-ISEP department in terms of head subjects, offshoots and gaps.



Mapping of DEI−ISEP Cluster 2 on ACM−CCS taxonomy

Figure 8: Mapping of DEI-ISEP cluster 2 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.



Mapping of DEI−ISEP Cluster 3 on ACM−CCS taxonomy

Figure 9: Mapping of DEI-ISEP cluster 3 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.15$.

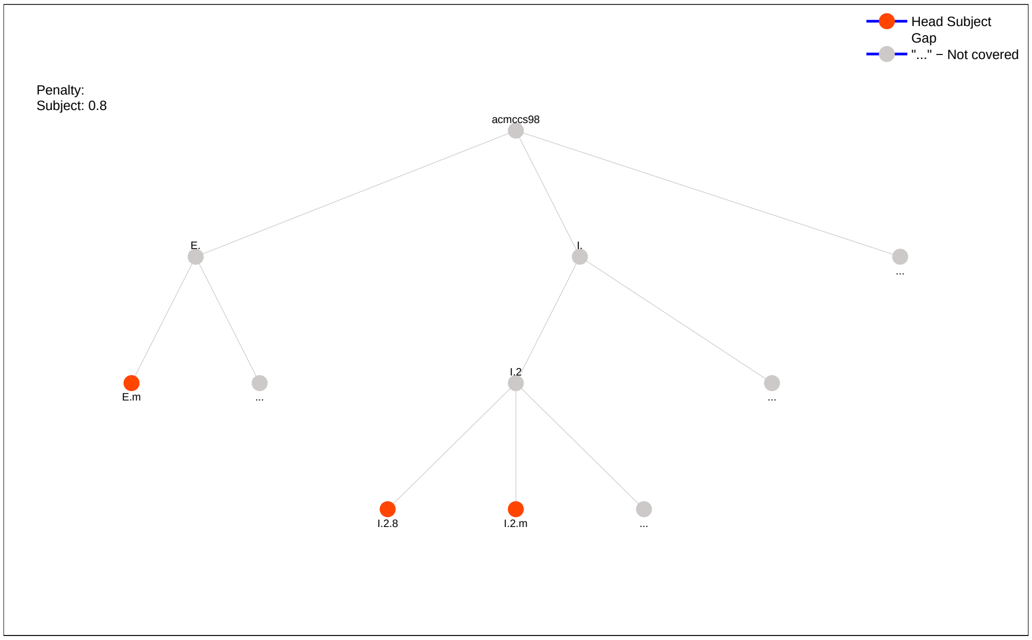


Figure 10: Mapping of DEI-ISEP cluster 4 onto the ACM-CCS tree with penalties $h = 1$, $o = 0.8$ and $g = 0.14$.

Table 13: Parsimonious Representation of DEI-ISEP Cluster 1

	HEAD SUBJECTS
D.2	SOFTWARE ENGINEERING
	OFFSHOTS
J.3	LIFE AND MEDICAL SCIENCES
K.7.3	Testing, Certification, and Licensing
	GAPS
D.2.0	General in D.2 - SOFTWARE ENGINEERING
D.2.3	Coding Tools and Techniques
D.2.4	Software/Program Verification
D.2.5	Testing and Debugging
D.2.6	Programming Environments
D.2.7	Distribution, Maintenance, and Enhancement
D.2.8	Metrics

Table 14: Parsimonious Representation of DEI-ISEP Cluster 2

	HEAD SUBJECTS
D.3	PROGRAMMING LANGUAGES
F.3	LOGICS AND MEANINGS OF PROGRAMS
	OFFSHOTS
C.2.4	Distributed Systems
D.1.6	Logic Programming
D.4.1	Process Management
H.3.4	Systems and Software
	GAPS
D.3.0	General in D.3 - PROGRAMMING LANGUAGES
D.3.2	Language Classifications
D.3.m	Miscellaneous in D.3 - PROGRAMMING LANGUAGES
F.3.0	General in F.3 - LOGICS AND MEANINGS OF PROGRAMS
F.3.1	Specifying and Verifying and Reasoning about Programs
F.3.m	Miscellaneous in F.3 - LOGICS AND MEANINGS OF PROGRAMS

Table 15: Parsimonious Representation of DEI-ISEP Cluster 3

	HEAD SUBJECTS
H.2	DATABASE MANAGEMENT
I.5	PATTERN RECOGNITION
	OFFSHOTS
H.3.3	Information Search and Retrieval
J.2	PHYSICAL SCIENCES AND ENGINEERING
	GAPS
H.2.0	General in H.2 - DATABASE MANAGEMENT
H.2.2	Physical Design
H.2.3	Languages
H.2.4	Systems
H.2.6	Database Machines
H.2.7	Database Administration
H.2.m	Miscellaneous in H.2 - DATABASE MANAGEMENT
I.5.0	General in I.5 - PATTERN RECOGNITION
I.5.1	Models
I.5.2	Design Methodology
I.5.m	Miscellaneous in I.5 - PATTERN RECOGNITION

Table 16: Parsimonious Representation of DEI-ISEP Cluster 4

	SUBJECTS
E.m	MISCELLANEOUS in E. - Data
I.2.8	Problem Solving, Control Methods, and Search
I.2.m	Miscellaneous in I.2 - ARTIFICIAL INTELLIGENCE

Table 17: Summary description of main Thematic Clusters

	N. of Thematic Clusters	HEAD SUBJECTS
CENTRIA-UNL	2	H. Information Systems I.5 PATTERN RECOGNITION G. Mathematics of Computing
CSIS-BKUL	4	D.2 SOFTWARE ENGINEERING J. Computer Applications E.2 DATA STORAGE REPRESENTATIONS H.0 GENERAL in H. - Information Systems I.0 GENERAL in I. - Computing Methodologies F. Theory of Computation I.2 ARTIFICIAL INTELLIGENCE I.5 PATTERN RECOGNITION
DEI-ISEP	4	D.2 SOFTWARE ENGINEERING D.3 PROGRAMMING LANGUAGES F.3 LOGICS AND MEANINGS OF PROGRAMS H.2 DATABASE MANAGEMENT I.5 PATTERN RECOGNITION E.m MISCELLANEOUS in E. - Data I.2.8 Problem Solving, Control Methods, and Search I.2.m Miscellaneous in I.2 - ARTIFICIAL INTELLIGENCE

5 Summary Description

In summary, referring to just head subjects obtained, without any references to gaps and offshoots, our results allow us to describe the departments' research, in most general terms, as summarized in Table 17.

6 Concluding Remarks

Since CENTRIA-UNL is a research center whereas CSIS-BKUL and DEI-ISEP are university departments, one should expect that the total number of research topics in CENTRIA-UNL is smaller than that in CSIS-BKUL or DEI-ISEP, and, similarly, the number of clusters in CENTRIA-UNL should be less than that in CSIS-BKUL or DEI-ISEP. Indeed, research centers are usually created for a limited set of research goals, whereas university departments must cover a wide range of topics in teaching, which necessarily affects the research efforts. The number of ACM-CCS topics scored in CENTRIA-UNL is 46 versus 54, 65 in CSIS-BKUL, DEI-ISEP, respectively. Also, the number of found clusters in CENTRIA-UNL is two, whereas in CSIS-BKUL and DEI-ISEP, it is four.

Overall, the found clustering results are consistent with the informal assessment of the research conducted in each of the research organizations. Moreover, the sets of research topics that have been chosen by individual members at the ESSA survey follow the cluster structure rather closely, falling mostly within one of them.

It should be noticed that the contribution weights of clusters reflect the tightness of clusters rather than their priority stand. Indeed, CENTRIA's two clusters contributions total of about 50%, CSIS-BKUL's four clusters contributions of 60%, and DEI-ISEP's four clusters contributions of 70%, which are good results for clustering.

Notice that depending on the relative weights of the events taken into account, the lifting can be done differently, leading to different portrayals of the cluster on ACM-CCS tree. In the present analysis, the penalties for "head subjects" (h), "offshoots" (o) and "gaps" g) of $h = 1$, $o = 0.8$, and $g = 0.15$, have been fixed empirically. However, the fine tunning of the penalty weights is an aspect to be addressed.

7 Acknowledgments

The participation of members of CENTRIA, Universidade Nova de Lisboa, Lisboa, Portugal, the Department of Computer Science and Information Systems, Birkbeck University of London, London UK, and DEI- ISEP, Porto, Portugal, in collecting the data on their research activities is gratefully acknowledge.

References

[1] B. Mirkin and S. Nascimento, "Analysis of Community Structure, Affinity Data and Research Activities using Additive Fuzzy Spectral Clustering", Technical Report 6, School of Computer Science, Birkbeck University of London, 2009.

[2] B. Mirkin, S. Nascimento, L.M. Pereira, "Cluster-lift method for mapping research activities over a concept tree", In: J. Koronacki, S.T. Wierzchon, Z.W. Ras, J. Kacprzyk (eds.), *Recent Advances in Machine Learning II– Dedicated to the memory of Ryszard S. Michalski, Computational Intelligence Series* Vol. 263, Springer, pp. 245-258, 2010.