

Studies in  
Informatics and Information Technologies

**Cognitive Traveling in Digital Space  
of the Web and Digital Libraries**

*Tradice* 



Pavol Návrat, Mária Bieliková  
Michal Laclavík, Ján Paralič  
Jela Steinerová (Eds.)

**COGNITIVE TRAVELING  
IN DIGITAL SPACE OF THE WEB  
AND DIGITAL LIBRARIES**

Yield of the Interdisciplinary Multi-Partner Project TraDice

*Slovak University of Technology in Bratislava  
Institute of Informatics, Slovak Academy of Sciences  
Technical University of Košice  
Comenius University in Bratislava*

Studies in Informatics and Information Technologies  
Cognitive Traveling in Digital Space of the Web and Digital Libraries

Editors:

*Pavol Návrat, Mária Bieliková*

Institute of Informatics and Software Engineering  
Faculty of Informatics  
and Information Technologies  
Slovak University of Technology in Bratislava  
Ilkovičova 2, 842 16 Bratislava, Slovakia  
{pavol.navrat,maria.bielikova}@stuba.sk

*Ján Paralič*

Department of Cybernetics  
and Artificial Intelligence  
Faculty of Electrical Engineering and Informatics  
Technical University of Košice  
Letná 9, 042 00 Košice, Slovakia  
{jan.paralic}@tuke.sk

*Michal Laclavík*

Institute of Informatics  
Slovak Academy of Science  
Dúbravská cesta 9, 845 07 Bratislava, Slovakia  
{michal.laclavik}@savba.sk

*Jela Steinerová*

Department of Library and Information Science  
Faculty of Philosophy  
Comenius University in Bratislava  
Šafárikovo nám. 6, 814 99 Bratislava, Slovakia  
{steinerova}@fphil.uniba.sk

© 2014, The authors mentioned in the Table of Contents

All contributions were reviewed by at least two reviewers. They appear here as submitted by the authors, possibly with only minor technical modifications.

This publication is a partial result of research project Cognitive travelling in digital space of the Web and digital libraries supported by personalized services and social networks supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10. Research in several themes was also partially supported by the Slovak Scientific Grant Agency under the contracts No. VEGA 1/0971/11 and VEGA 1/0675/11.

Production of this publication was partially supported also by the Gratex IT Institute, within the framework of the GraFIIT Foundation ([www.gratex.com](http://www.gratex.com))

Visit TraDice on the Web: [tradice.fiit.stuba.sk](http://tradice.fiit.stuba.sk)

Executive editors: Michal Kompan, Alena Kovárová

Published by STU Press

Nakladateľstvo STU, Vazovova 5, Bratislava, Slovakia

ISBN print 978-80-227-4254-2

ISBN online 978-80-227-4263-4

## PREFACE

This book summarizes yield of the project TraDice. Four institutions from three universities and one research institute, from two cities, successfully applied and received support from the Slovak Research and Development Agency for their research project Cognitive travelling in digital space of the web and digital libraries supported by personalized services and social networks. The project ran from 2011 till 2014.

The research theme is quite faithfully described by the title of the project. We ventured to investigate ways and means of seeking information in the information space formed by the Web and by digital libraries. Exploratory search has been just one of the foundation concepts of our research. Semantic web, personalization, context have been among the other crucial concepts influencing our research approach.

The research has been devised as interdisciplinary. This reflects the very nature of the problem area. To investigate information behavior of a typical user or a group of them, a more comprehensive view is necessary. For example, the concept of a user is too generic. We have to identify her specific role in the process of using search tools. She has some information need. Simply put, she is curious. She is interested in something. But such an interested fellow does not act in isolation. She is always a social actor, too. Currently, there are emerging various web based social networking services that facilitate sharing, exchanging or recommending of information. If interested fellows are willing to invest some effort in commenting, annotating, tagging what they read, they can expect more appropriate answers or advices in return, provided intelligent tools are designed, utilizing such concepts as e.g., wisdom of the crowd.

The project team comprises of members from four distinct institutes. The Institute of Informatics and Software Engineering from the Slovak University of Technology in Bratislava has assumed the leading role in the project from the very beginning. They invited the other three institutes to write the project proposal: Department of Cybernetics and Artificial Intelligence from the Technical University of Kosice, Department of Library and Information Science from the Comenius University in Bratislava, and the Institute of Informatics of the Slovak Academy of Sciences in Bratislava. From that moment on, however, there were four equal partners on the board. We wrote the project proposal together and, when the grant was awarded, conducted the research in collaboration.

One of the important vehicles driving collaboration were project workshops. They were organized annually within the framework of the Workshop

on Intelligent and Knowledge Oriented Technologies WIKT, held in Herľany in 2011, in Smolenice in 2012 and in Herľany in 2013 again. Project after-meeting was in Smolenice in 2014. We wish to express our gratitude to the organizers of these events for their willingness to host our project meetings.

This book has been first published in its preliminary version on the web. It comprises 48 original contributions describing results of research within this project. The contributions are of two kinds, reflecting the reality of our publishing activities. Most of the results that have been achieved in this project, have also already been published in various journals or conference proceedings. In such cases, this book includes a paper with a brief description of the motivation and results. Important thing is that the references include an explicit reference to the extensive publication(s) of these results published elsewhere. We have 44 such brief papers referring to 118 papers authored by project researchers but published elsewhere. Another kind of contribution presents original results not yet published elsewhere. Now when the project has been completed and all the contributions have been gathered, the final version of the book is published in print.

The volume could not be produced without an extraordinary and genuine effort of many people who all deserve appreciation. First, I should like to mention my fellow co-editors who supported the idea throughout the whole production process. Perhaps the greatest thanks should go to the authors of contributed papers. But I should like to express my gratitude also to executive editors Michal Kompan and Alena Kovárová who worked tirelessly, reliably and creatively.

We hope the volume presents interesting research results that will attract attention of the related research community. We desire it will also serve as a vehicle for dissemination of the project's results.

Pavol Návrat

# CONTENTS

(in the order of submission date)

Cognitive Traveling in Digital Space <i>Pavol Návrat</i> .....	1
Identification and Tracking of News Stories Using Honey Bee Inspired Agents <i>Štefan Sabo</i> .....	4
Context and Group Modelling to Better Recommendations <i>Michal Kompan, Mária Bieliková</i> .....	6
Lightweight Semantics Modeling and Acquisition for the “Wild Web” <i>Mária Bieliková, Michal Barla, Marián Šimko</i> .....	10
Wikipedia Explicit Semantics Based Entity Disambiguation <i>Jozef Tvarožek</i> .....	14
Multimedia Archive Organization and Domain Modelling Using GWAPs <i>Jakub Šimko, Mária Bieliková</i> .....	17
Semantization of Research Group's Wiki <i>Jakub Šimko, Martin Markech</i> .....	21
Process-oriented Analysis of Web-based Data <i>František Babič, Ján Paralič</i> .....	27
Adaptive Query Formulation Improvement <i>Ladislav Martinský, Pavol Návrat</i> .....	32
Arranging Face-To-Face Events Using Social Networking Services <i>Jaroslav Michalco, Pavol Návrat</i> .....	36
Exploratory Search on Social Networking Sites: How User Feedback and Multi-perspective Microblog Analysis Can Be Employed on Twitter <i>Michal Žilinčík, Pavol Návrat, Gabriela Kosková</i> .....	40
Blog Clustering Enhanced by Mining the Web Comments <i>Tomáš Kuzár, Pavol Návrat</i> .....	43
Learning to Rank Scientific Papers <i>Tomáš Kramár, Martin Petluš</i> .....	46
Modeling the Dynamics of Research Interests <i>Tomáš Kramár, Roman Bilevic</i> .....	50

Detecting Search Sessions Using Document Metadata and Implicit Feedback. <i>Tomáš Kramár, Mária Bieliková</i> .....	56
PeWeProxy: a Platform for Ubiquitous Web Adaptation and User Modeling. <i>Tomáš Kramár, Michal Barla, Mária Bieliková</i> .....	58
Evaluation Framework for Personalized Recommendations <i>Jakub Ševcech, Michal Kompan, Mária Bieliková</i> .....	61
Vector-Based Tree News Recommendation <i>Michal Kompan, Dušan Zeleník, Mária Bieliková</i> .....	63
Annotations as User Interest Indicators for Related Document Retrieval <i>Jakub Ševcech, Mária Bieliková</i> .....	67
Cluster-based and Graph-based Browsing and Exploring Search Results <i>Karol Rástočný, Mária Bieliková</i> .....	71
On Different Aspects of Usage of One-sided Concept Lattices in Data Analysis <i>Peter Butka</i> .....	74
Bee Hive at Work: Model with Searching and Optimizing Potential <i>Pavol Návrát, Anna Bou Ezzeddine</i> .....	78
Automatic Image Annotation Based on Visual Content Analysis <i>Eduard Kuric, Mária Bieliková</i> .....	82
Utilizing History of Navigation in a Digital Library for Term Cloud Personalization <i>Samuel Molnár, Róbert Móro, Mária Bieliková</i> .....	86
The Role of Text Summarization in Digital Libraries <i>Róbert Móro, Mária Bieliková</i> .....	89
Effective Corpora Creation for Sentiment Analysis <i>Peter Koncz, Ján Paralič</i> .....	92
Identifying Semantic Relationships in Digital Libraries <i>Zoltán Harsányi, Viera Rozinajová</i> .....	95
Automatic Annotation of Non-English Web Content <i>Jakub Ševcech, Mária Bieliková</i> .....	99
Browser-based User Interest Modelling <i>Márius Šajgalík, Michal Barla, Mária Bieliková</i> .....	102
Domain Modelling in Socially Aware Information Spaces <i>Marián Šimko</i> .....	106
Advanced Information Sources for User Modelling <i>Marián Šimko</i> .....	109
Human Behavior as an Aspect for Context Inference <i>Dušan Zeleník, Mária Bieliková</i> .....	111



Recommendation Based on Graph Algorithms <i>Dušan Zeleník, Mária Bielíková</i> .....	115
Utilization of Behavior of Users for Personalized Link Recommendation <i>Michal Holub, Mária Bielíková</i> .....	119
Combining Named Entity Recognition Tools <i>Štefan Dlugolinský, Peter Krammer, Marek Ciglan, Michal Laclavík, Ladislav Hluchý</i> .....	121
Entity Search in Text Graphs <i>Michal Laclavík, Štefan Dlugolinský, Marek Ciglan, Martin Šeleng, Ladislav Hluchý</i> .....	125
Text Clustering on Web with Frequent Itemsets - FICWAN <i>Tomáš Kučečka, Daniela Chudá</i> .....	128
Personalised Recommendations for Group of Users <i>Michal Kompan, Mária Bielíková</i> .....	130
Adaptive Support for Collaborative Knowledge Sharing <i>Ivan Srba, Mária Bielíková</i> .....	134
Exploration and Organization of Digital Library Information Space <i>Róbert Móro, Mária Bielíková</i> .....	137
Annota: Platform for Conducting Experiments in Digital Libraries Information Space <i>Róbert Móro, Michal Holub, Jakub Ševcech, Mária Bielíková</i> .....	141
Mining Challenges on Social Networks Data <i>Gabriel Tutoky, Ján Paralič</i> .....	145
Some Results Related to Theory and Practice of Fuzzy Concept Lattices <i>Peter Butka</i> .....	148
Leveraging Feature Vectors to Extract Discriminative Words <i>Márius Šajgalík, Michal Barla, Mária Bielíková</i> .....	152
Tabbed Browsing Behaviour in Digital Space <i>Martin Labaj, Mária Bielíková</i> .....	155
Ecological Information Interactions for Digital Scholarship <i>Jela Steinerová</i> .....	158
Search Query Expansion Based on Searcher's Activity Context <i>Lubomír Vnenk, Mária Bielíková</i> .....	170
Visualizing Multiple News Stories Using the New Galactic and Concurrent Stream Meta- phors <i>Alena Kovárová</i> .....	173
Author Index.....	177



# CONTENTS

Cognitive Traveling in Digital Space <i>Pavol Návrat</i> .....	1
---	---

## Section I: Exploration and Navigation in Digital Space

---

Identification and Tracking of News Stories Using Honey Bee Inspired Agents <i>Štefan Sabo</i> .....	4
Exploratory Search on Social Networking Sites: How User Feedback and Multi-perspective Microblog Analysis Can Be Employed on Twitter <i>Michal Žilínčik, Pavol Návrat, Gabriela Kosková</i> .....	40
Cluster-based and Graph-based Browsing and Exploring Search Results <i>Karol Rástočný, Mária Bieliková</i> .....	71
Bee Hive at Work: Model with Searching and Optimizing Potential <i>Pavol Návrat, Anna Bou Ezzeddine</i> .....	78
Utilizing History of Navigation in a Digital Library for Term Cloud Personalization <i>Samuel Molnár, Róbert Móro, Mária Bieliková</i> .....	86
Exploration and Organization of Digital Library Information Space <i>Róbert Móro, Mária Bieliková</i> .....	137
Annota: Platform for Conducting Experiments in Digital Libraries Information Space <i>Róbert Móro, Michal Holub, Jakub Ševcech, Mária Bieliková</i> .....	141
Visualizing Multiple News Stories Using the New Galactic and Concurrent Stream Metaphors <i>Alena Kovárová</i> .....	173

## Section II: Search and Recommendation in Digital Space

---

Adaptive Query Formulation Improvement <i>Ladislav Martinský, Pavol Návrat</i> .....	32
Arranging Face-To-Face Events Using Social Networking Services <i>Jaroslav Michalco, Pavol Návrat</i> .....	36

Learning to Rank Scientific Papers <i>Tomáš Kramár, Martin Petluš</i> .....	46
Detecting Search Sessions Using Document Metadata and Implicit Feedback. <i>Tomáš Kramár, Mária Bielíková</i> .....	56
Evaluation Framework for Personalized Recommendations <i>Jakub Ševcech, Michal Kompan, Mária Bielíková</i> .....	61
Vector-Based Tree News Recommendation <i>Michal Kompan, Dušan Zeleník, Mária Bielíková</i> .....	63
Annotations as User Interest Indicators for Related Document Retrieval <i>Jakub Ševcech, Mária Bielíková</i> .....	67
Recommendation Based on Graph Algorithms <i>Dušan Zeleník, Mária Bielíková</i> .....	115
Entity Search in Text Graphs <i>Michal Laclavík, Štefan Dlugolinský, Marek Ciglan, Martin Šeleng, Ladislav Hluchý</i> .....	125
Personalised Recommendations for Group of Users <i>Michal Kompan, Mária Bielíková</i> .....	130
Adaptive Support for Collaborative Knowledge Sharing <i>Ivan Srba, Mária Bielíková</i> .....	134
Search Query Expansion Based on Searcher's Activity Context <i>Lubomír Vnenk, Mária Bielíková</i> .....	170

### **Section III: User, Context and Group Modelling**

---

Context and Group Modelling to Better Recommendations <i>Michal Kompan, Mária Bielíková</i> .....	6
Modeling the Dynamics of Research Interests <i>Tomáš Kramár, Roman Bilevic</i> .....	50
PeWeProxy: a Platform for Ubiquitous Web Adaptation and User Modeling. <i>Tomáš Kramár, Michal Barla, Mária Bielíková</i> .....	58
Browser-based User Interest Modelling <i>Márius Šajgalík, Michal Barla, Mária Bielíková</i> .....	102
Advanced Information Sources for User Modelling <i>Marián Šimko</i> .....	109
Human Behavior as an Aspect for Context Inference <i>Dušan Zeleník, Mária Bielíková</i> .....	111
Utilization of Behavior of Users for Personalized Link Recommendation <i>Michal Holub, Mária Bielíková</i> .....	119

Ecological Information Interactions for Digital Scholarship <i>Jela Steinerová</i> .....	158
---	-----

#### **Section IV: Domain Modelling**

---

Lightweight Semantics Modeling and Acquisition for the “Wild Web” <i>Mária Bieliková, Michal Barla, Marián Šimko</i> .....	10
Wikipedia Explicit Semantics Based Entity Disambiguation <i>Jozef Tvarožek</i> .....	14
Multimedia Archive Organization and Domain Modelling Using GWAPs <i>Jakub Šimko, Mária Bieliková</i> .....	17
Semantization of Research Group’s Wiki <i>Jakub Šimko, Martin Markech</i> .....	21
Automatic Image Annotation Based on Visual Content Analysis <i>Eduard Kuric, Mária Bieliková</i> .....	82
The Role of Text Summarization in Digital Libraries <i>Róbert Móro, Mária Bieliková</i> .....	89
Identifying Semantic Relationships in Digital Libraries <i>Zoltán Harsányi, Viera Rozinajová</i> .....	95
Automatic Annotation of Non-English Web Content <i>Jakub Ševcech, Mária Bieliková</i> .....	99
Domain Modelling in Socially Aware Information Spaces <i>Marián Šimko</i> .....	106
Mining Challenges on Social Networks Data <i>Gabriel Tutoky, Ján Paralič</i> .....	145
Some Results Related to Theory and Practice of Fuzzy Concept Lattices <i>Peter Butka</i> .....	148
Leveraging Feature Vectors to Extract Discriminative Words <i>Márius Šajgalík, Michal Barla, Mária Bieliková</i> .....	152
Tabbed Browsing Behaviour in Digital Space <i>Martin Labaj, Mária Bieliková</i> .....	155

#### **Section V: Data Analysis, Data Mining and Knowledge Discovery**

---

Process-oriented Analysis of Web-based Data <i>František Babič, Ján Paralič</i> .....	27
--	----

Blog Clustering Enhanced by Mining the Web Comments <i>Tomáš Kuzár, Pavol Návrat</i> .....	43
On Different Aspects of Usage of One-sided Concept Lattices in Data Analysis <i>Peter Butka</i> .....	74
Effective Corpora Creation for Sentiment Analysis <i>Peter Koncz, Ján Paralič</i> .....	92
Combining Named Entity Recognition Tools <i>Štefan Dlugolinský, Peter Krammer, Marek Ciglan, Michal Laclavík, Ladislav Hluchý</i> .....	121
Text Clustering on Web with Frequent Itemsets - FICWAN <i>Tomáš Kučečka, Daniela Chudá</i> .....	128