

Dmitry Ignatov, Sergei Kuznetsov, Jonas Poelmans (Eds.)

## **CDUD'11 – Concept Discovery in Unstructured Data**

Workshop co-located with the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC-2011)  
June 2011, Moscow, Russia

The proceedings are published online in the CEUR-Workshop series (ISSN 1613-0073) and the volume Vol-757 has a unique URN: urn:nbn:de:0074-757-4.

## **Volume Editors**

Dmitry Ignatov  
School of Applied Mathematics and Informatics  
National Research University Higher School of Economics, Moscow, Russia

Sergei Kuznetsov  
School of Applied Mathematics and Informatics  
National Research University Higher School of Economics, Moscow, Russia

Jonas Poelmans  
Faculty of Business and Economics  
Katholieke Universiteit Leuven, Belgium

Copyright © 2011 for the individual papers by papers' authors, for the Volume by the editors. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means without the prior permission of the copyright owners.

## Preface

Concept discovery is a Knowledge Discovery in Databases (KDD) research field that uses human-centered techniques such as Formal Concept Analysis (FCA), Biclustering, Triclustering, Conceptual Graphs etc. for gaining insight into the underlying conceptual structure of the data. Traditional machine learning techniques are mainly focusing on structured data whereas most data available resides in unstructured, often textual, form. Compared to traditional data mining techniques, human-centered instruments actively engage the domain expert in the discovery process.

This volume contains the contributions to CDUD 2011, the International Workshop on Concept Discovery in Unstructured Data (CDUD) held in Moscow. The main goal of this workshop was to provide a forum for researchers and developers of data mining instruments working on issues with analyzing unstructured data.

We are proud that we could welcome 13 valuable contributions to this volume. The majority of the accepted papers described innovative research on data discovery in unstructured texts. Authors worked on issues such as transforming unstructured into structured information by amongst others extracting keywords and opinion words from texts with Natural Language Processing methods. Multiple authors who participated in the workshop used methods from the conceptual structures field including Formal Concept Analysis and Conceptual Graphs. Applications include but are not limited to text mining police reports, sociological definitions, movie reviews, etc.

Last but not least, we would like to thank the administration of the Higher School of Economics who took care of all arrangements to make this conference pleasant and enjoyable.

June 2011, Moscow

Dmitry Ignatov  
Sergei Kuznetsov  
Jonas Poelmans

# Organization

This CDUD'11 workshop was held in June 2011 in Moscow, Russia co-located with the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC-2011) at the National Research University Higher School of Economics.

## Program Chairs

Dmitry Ignatov	State University Higher School of Economics, Russia
Sergei Kuznetsov	State University Higher School of Economics, Russia
Jonas Poelmans	Katholieke Universiteit Leuven, Belgium

## Program Committee

Guido Dedene	Katholieke Universiteit Leuven, Belgium
	Amsterdam Business School, The Netherlands
Paul Elzinga	Amsterdam-Amstelland Police, The Netherlands
Bernhard Ganter	Dresden University of Technology, Germany
Richard Hill	University of Derby, UK
Alex Neznanov	State University Higher School of Economics, Russia
Simon Polovina	University of Sheffield, UK
Henrik Scharfe	Aalborg University, Denmark
Vladimir Selegey	ABBYY, Russia
Stijn Viaene	Katholieke Universiteit Leuven, Belgium
Laszlo Szathmary	University of Quebec in Montreal, Canada

## Sponsoring Institutions

ABBYY, Moscow  
Russian Foundation for Basic Research, Moscow  
Poncelet Laboratory (UMI 2615 du CNRS), Moscow  
State University Higher School of Economics, Moscow  
Yandex, Moscow  
Witology, Moscow  
Dynasty Foundation, Moscow

## Table of Contents

Automatic Entity Detection Based on News Cluster Structure . . . . .	1
<i>Aleksey Alekseev and Natalia Loukachevitch</i>	
Application of Conceptual Structures in Requirements Modeling . . . . .	11
<i>Michael Bogatyrev and Vadim Nuriahmetov</i>	
Abstracting Concepts from Text Documents by Using an Ontology . . . . .	21
<i>Ekaterina Cherniak, Olga Chugunova, Julia Askarova, Susana Nascimento and Boris Mirkin</i>	
Extraction and Use of Opinion Words for Three-Way Review Classification Task . . . . .	31
<i>Ilia Chetviorkin and Natalia Loukachevitch</i>	
Constructing Galois Lattice in Good Classification Tests Mining . . . . .	43
<i>Xenia Naidenova</i>	
Concept Relation Discovery and Innovation Enabling Technology (CORDIET) . . . . .	53
<i>Jonas Poelmans, Paul Elzinga, Alexey Neznanov, Stijn Viaene, Sergei Kuznetsov, Dmitry Ignatov and Guido Dedene</i>	
Concept Lattice Implementation in Semantic Structuring of Adjectives . .	63
<i>Serge Potemkin</i>	
Exploring Semantic Orientation of Adverbs . . . . .	71
<i>Serge Potemkin and Galina Kedrova</i>	
The Third Personal Pronoun Anaphora Resolution in Texts from Narrow Subject Domains with Grammatical Errors and Mistypings . . . . .	79
<i>Daniel Skatov and Sergey Liverko</i>	
An FCA-Based Approach to the Study of Socialization Definitions . . . . .	93
<i>Sergei Vinkov</i>	
Temporal Concept Analysis Explained by Examples . . . . .	104
<i>Karl Erich Wolff</i>	
Research Challenges of Dynamic Socio-Semantic Networks . . . . .	119
<i>Rostislav Yavorsky</i>	
Recommender System Based on Algorithm of Bicluster Analysis RecBi . .	122
<i>Dmitry Ignatov, Jonas Poelmans and Vasily Zaharchuk</i>	