

Josiane Mothe · Jacques Savoy  
Jaap Kamps · Karen Pinel-Sauvagnat  
Gareth J.F. Jones · Eric SanJuan  
Linda Cappellato · Nicola Ferro (Eds.)

# Experimental IR Meets Multilinguality, Multimodality, and Interaction

6th International Conference  
of the CLEF Association, CLEF'15  
Toulouse, France, September 8–11, 2015  
Proceedings

# Contents

## Experimental IR

Experimental Study on Semi-structured Peer-to-Peer Information Retrieval Network . . . . .	3
<i>Rami S. Alkhaldeh and Joemon M. Jose</i>	
Evaluating Stacked Marginalised Denoising Autoencoders Within Domain Adaptation Methods . . . . .	15
<i>Boris Chidlovskii, Gabriela Csurka, and Stephane Clinchant</i>	
Language Variety Identification Using Distributed Representations of Words and Documents . . . . .	28
<i>Marc Franco-Salvador, Francisco Rangel, Paolo Rosso, Mariona Taulé, and M. Antònia Martí</i>	
Evaluating User Image Tagging Credibility . . . . .	41
<i>Alexandru Lucian Ginsca, Adrian Popescu, Mihai Lupu, Adrian Iftene, and Ioannis Kanellos</i>	

## Web and Social Media

Tweet Expansion Method for Filtering Task in Twitter . . . . .	55
<i>Payam Karisani, Farhad Oroumchian, and Maseud Rahgozar</i>	
Real-Time Entity-Based Event Detection for Twitter . . . . .	65
<i>Andrew J. McMinn and Joemon M. Jose</i>	
A Comparative Study of Click Models for Web Search . . . . .	78
<i>Artem Grotov, Aleksandr Chuklin, Ilya Markov, Luka Stout, Finde Xumara, and Maarten de Rijke</i>	
Evaluation of Pseudo Relevance Feedback Techniques for Cross Vertical Aggregated Search . . . . .	91
<i>Hermann Ziak and Roman Kern</i>	

## Long Papers with Short Presentation

Analysing the Role of Representation Choices in Portuguese Relation Extraction . . . . .	105
<i>Sandra Collovini, Marcelo de Baires P. Filho, and Renata Vieira</i>	

An Investigation of Cross-Language Information Retrieval for User-Generated Internet Video. . . . .	117
<i>Ahmad Khwileh, Debasis Ganguly, and Gareth J.F. Jones</i>	
Benchmark of Rule-Based Classifiers in the News Recommendation Task . . .	130
<i>Tomáš Kliegr and Jaroslav Kuchař</i>	
Enhancing Medical Information Retrieval by Exploiting a Content-Based Recommender Method. . . . .	142
<i>Wei Li and Gareth J.F. Jones</i>	
Summarizing Citation Contexts of Scientific Publications. . . . .	154
<i>Sandra Mitrović and Henning Müller</i>	
A Multiple-Stage Approach to Re-ranking Medical Documents. . . . .	166
<i>Heung-Seon Oh, Yuchul Jung, and Kwang-Young Kim</i>	
Exploring Behavioral Dimensions in Session Effectiveness. . . . .	178
<i>Teemu Pääkkönen, Kalervo Järvelin, Jaana Kekäläinen, Heikki Keskustalo, Feza Baskaya, David Maxwell, and Leif Azzopardi</i>	
<b>Short Papers</b>	
META TEXT ALIGNER: Text Alignment Based on Predicted Plagiarism Relation. . . . .	193
<i>Samira Abnar, Mostafa Dehghani, and Azadeh Shakery</i>	
Automatic Indexing of Journal Abstracts with Latent Semantic Analysis . . . .	200
<i>Joel Robert Adams and Steven Bedrick</i>	
Shadow Answers as an Intermediary in Email Answer Retrieval . . . . .	209
<i>Alyaa Alfalahi, Gunnar Eriksson, and Eriks Sneiders</i>	
Are Topically Diverse Documents Also Interesting? . . . . .	215
<i>Hosein Azarbonyad, Ferron Saan, Mostafa Dehghani, Maarten Marx, and Jaap Kamps</i>	
Modeling of the Question Answering Task in the YodaQA System. . . . .	222
<i>Petr Baudiš and Jan Šedivý</i>	
Unfair Means: Use Cases Beyond Plagiarism . . . . .	229
<i>Paul Clough, Peter Willett, and Jessie Lim</i>	
Instance-Based Learning for Tweet Monitoring and Categorization . . . . .	235
<i>Julien Gobeill, Arnaud Gaudinat, and Patrick Ruch</i>	

Are Test Collections “Real”? Mirroring Real-World Complexity  
in IR Test Collections . . . . . 241  
*Melanie Imhof and Martin Braschler*

Evaluation of Manual Query Expansion Rules on a Domain Specific  
FAQ Collection . . . . . 248  
*Mladen Karan and Jan Šnajder*

Evaluating Learning Language Representations . . . . . 254  
*Jussi Karlgren, Jimmy Callin, Kevyn Collins-Thompson,  
Amaru Cuba Gyllensten, Ariel Ekgren, David Jurgens, Anna Korhonen,  
Fredrik Olsson, Magnus Sahlgren, and Hinrich Schütze*

Automatic Segmentation and Deep Learning of Bird Sounds . . . . . 261  
*Hendrik Vincent Koops, Jan van Balen, and Frans Wiering*

The Impact of Noise in Web Genre Identification . . . . . 268  
*Dimitrios Pritsos and Efstathios Stamatas*

On the Multilingual and Genre Robustness of EmoGraphs for Author  
Profiling in Social Media . . . . . 274  
*Francisco Rangel and Paolo Rosso*

Is Concept Mapping Useful for Biomedical Information Retrieval? . . . . . 281  
*Wei Shen and Jian-Yun Nie*

Using Health Statistics to Improve Medical and Health Search . . . . . 287  
*Tawan Sierek and Allan Hanbury*

Determining Window Size from Plagiarism Corpus for Stylometric Features . . . 293  
*Šimon Suchomel and Michal Brandejs*

Effect of Log-Based Query Term Expansion on Retrieval Effectiveness  
in Patent Searching . . . . . 300  
*Wolfgang Tannebaum, Parvaz Mahdabi, and Andreas Rauber*

Integrating Mixed-Methods for Evaluating Information Access Systems . . . . . 306  
*Simon Wakeling and Paul Clough*

Teaching the IR Process Using Real Experiments Supported by Game  
Mechanics . . . . . 312  
*Thomas Wilhelm-Stein and Maximilian Eibl*

Tweet Contextualization Using Association Rules Mining and DBpedia . . . . . 318  
*Meriem Amina Zingla, Chiraz Latiri, and Yahya Slimani*

**Best of the Labs**

Search-Based Image Annotation: Extracting Semantics from Similar Images. . . . .	327
<i>Petra Budikova, Michal Batko, Jan Botorek, and Pavel Zezula</i>	
NLP-Based Classifiers to Generalize Expert Assessments in E-Reputation . . .	340
<i>Jean-Valère Cossu, Emmanuel Ferreira, Killian Janod, Julien Gaillard, and Marc El-Bèze</i>	
A Method for Short Message Contextualization: Experiments at CLEF/INEX . . . . .	352
<i>Liana Ermakova</i>	
Towards Automatic Large-Scale Identification of Birds in Audio Recordings. . . . .	364
<i>Mario Lasseck</i>	
Optimizing and Evaluating Stream-Based News Recommendation Algorithms. . . . .	376
<i>Andreas Lommatzsch and Sebastian Werner</i>	
Information Extraction from Clinical Documents: Towards Disease/Disorder Template Filling. . . . .	389
<i>Veera Raghavendra Chikka, Nestor Mariyasagayam, Yoshiki Niwa, and Kamalakar Karlapalem</i>	
Adaptive Algorithm for Plagiarism Detection: The Best-Performing Approach at PAN 2014 Text Alignment Competition . . . . .	402
<i>Miguel A. Sanchez-Perez, Alexander Gelbukh, and Grigori Sidorov</i>	
Question Answering via Phrasal Semantic Parsing. . . . .	414
<i>Kun Xu, Yansong Feng, Songfang Huang, and Dongyan Zhao</i>	

**Labs Overviews**

Overview of the CLEF eHealth Evaluation Lab 2015 . . . . .	429
<i>Lorraine Goeuriot, Liadh Kelly, Hanna Suominen, Leif Hanlen, Aurélie Névéol, Cyril Grouin, João Palotti, and Guido Zuccon</i>	
General Overview of ImageCLEF at the CLEF 2015 Labs . . . . .	444
<i>Mauricio Villegas, Henning Müller, Andrew Gilbert, Luca Piras, Josiah Wang, Krystian Mikolajczyk, Alba G. Seco de Herrera, Stefano Bromuri, M. Ashraf ul Amin, Mahmood Kazi Mohammed, Burak Acar, Suzan Uskudarli, Neda B. Marvasti, José F. Aldana, and María del Mar Roldán García</i>	

LifeCLEF 2015: Multimedia Life Species Identification Challenges . . . . .	462
<i>Alexis Joly, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Pierre Bonnet, Willem-Pier Vellinga, Robert Planqué, Andreas Rauber, Simone Palazzo, Bob Fisher, and Henning Müller</i>	
Overview of the Living Labs for Information Retrieval Evaluation (LL4IR) CLEF Lab 2015 . . . . .	484
<i>Anne Schuth, Krisztian Balog, and Liadh Kelly</i>	
Stream-Based Recommendations: Online and Offline Evaluation as a Service . . . . .	497
<i>Benjamin Kille, Andreas Lommatzsch, Roberto Turrin, András Serény, Martha Larson, Torben Brodt, Jonas Seiler, and Frank Hopfgartner</i>	
Overview of the PAN/CLEF 2015 Evaluation Lab . . . . .	518
<i>Efstathios Stamatatos, Martin Potthast, Francisco Rangel, Paolo Rosso, and Benno Stein</i>	
Overview of the CLEF Question Answering Track 2015 . . . . .	539
<i>Anselmo Peñas, Christina Unger, Georgios Paliouras, and Ioannis Kakadiaris</i>	
Overview of the CLEF 2015 Social Book Search Lab . . . . .	545
<i>Marijn Koolen, Toine Bogers, Maria Gäde, Mark Hall, Hugo Huurdeman, Jaap Kamps, Mette Skov, Elaine Toms, and David Walsh</i>	
<b>Author Index . . . . .</b>	<b>565</b>