CLEF 2010

Labs and Workshops

22-23 September 2010, University of Padua, Italy

CLEF 2010 Padua

Abstracts

Notebook Papers 2010

Editors: Martin Braschler, Donna Harman, Emanuele Pianta

promise

http://www.clef2010.org

ISSN: 2038-4726
ISBN: 978-88-904810-0-0

The **CLEF 2010 Labs** continue the CLEF tradition of community-based benchmarking and complement it with workshops on emerging issues in evaluation methodology.

CLEF 2010 Abstract - Notebook Papers

22-23 September, University of Padua, Italy

CLEF-IP
A benchmarking activity on intellectual property

ImageCLEF
A benchmarking activity on image retrieval

PAN
A benchmarking activity on plagiarism and Wikipedia vandalism detection

RespubliQA
A benchmarking activity on question answering using multilingual political data

WePS
A benchmarking activity on Web people search

CriES
A workshop aimed at exploring the evaluation of searching for expertise in social media

LogCLEF
A workshop aimed at exploring methodologies for studying search engine log files

ISBN 978-88-904810-0-0

http://www.clef2010.org

Table of Contents

Introduction to the CLEF 2010 Labs
Martin Braschler and Donna Harman 15

Intellectual Property (CLEF-IP)

CLEF-IP 2010: Retrieval Experiments in the Intellectual Property Domain
Florina Piroi 17

CLEF-IP: Prior Art Candidate Search Task

Applying the KISS Principle for the CLEF- IP 2010 Prior Art Candidate Patent Search Task
Walid Magdy and Gareth J.F. Jones 18

Using BM25F and KLD for Pattern Retrieval
Joaquín Pérez-Iglesias, Álvaro Rodrigo, and Víctor Fresno 19

Prior Art Retrieval Using Various Patent Document Fields Contents
Metti Zakaria Wanagiri and Mirna Adriani 19

CLEF-IP 2010: Prior Art Retrieval Using the Different Sections in Patent Documents
Eva D'hondt and Suzan Verberne 20

Phrases or Terms? The Impact of Different Query Types
Daniela Becks, Thomas Mandl, and Christa Womser-Hacker 21

Building Strategies, a Year Later
Wouter Alink, Roberto Cornacchia, and Arjen P. de Vries 21

Experiments with Citation Mining and Key-Term Extraction for Prior Art Search
Patrice Lopez and Laurent Romary 22

Automatic Prior Art Searching and Patent Encoding at CLEF-IP '10
Douglas Teodoro, Julien Gobeill, Emilie Pasche, Dina Vishnyakova, Patrick Ruch, and Christian Lovis 22

CLEF-IP: Classification Task

LCI-INSA Linguistic Experiment for CLEF-IP Classification Track
Jean Beney 23

myClass: A Mature Tool for Patent Classification
Jacques Guyot, Karim Benzineb, and Gilles Falquet 24

Combining Semantics and Statistics for Patent Classification
Franck Derieux, Mihaela Bobeica, Delphine Pois, and Jean-Pierre Raysz 25

Patent Classification Experiments with the Linguistic Classification System LCS
Suzan Verberne, Merijn Vogel, and Eva D'hondt 25

Cross Language Image Retrieval (ImageCLEF)

Overview of the Wikipedia Retrieval Task at ImageCLEF 2010
Adrian Popescu, Theodora Tsikrika, and Jana Kludas 26

Overview of the CLEF 2010 Medical Image Retrieval Track
Henning Müller, Jayashree Kalpathy-Cramer, Ivan Eggel, Steven Bedrick, Joe Reisetter, Charles E. Kahn Jr., and William Hersh 27

The Robot Vision Track at ImageCLEF 2010
Andrzej Pronobis, Marco Fornoni, Henrik I. Christensen, and Barbara Caputo 28

New Strategies for Image Annotation: Overview of the Photo Annotation Task at ImageCLEF 2010
Stefanie Nowak and Mark Huiskes 28

ImageCLEF: Wikipedia Retrieval

Télécom Bretagne at ImageCLEF WikipediaMM 2010
Adrian Popescu 29

XRCE's Participation in Wikipedia Retrieval, Medical Image Modality Classification and Ad-hoc Retrieval Tasks of ImageCLEF 2010
Stéphane Clinchant, Gabriela Csurka, Julien Ah-Pine, Guillaume Jacquet, Florent Perronnin, Jorge Sánchez, and Keyvan Minoukadeh 30

Blind Relevance Feedback for the ImageCLEF Wikipedia Retrieval Task
Ray R. Larson 30

Document Expansion for Text-based Image Retrieval at WikipediaMM 2010
Jinming Min, Johannes Leveling, and Gareth J.F. Jones 31

Experiences at ImageCLEF 2010 using CBIR and TBIR Mixing Information Approaches
Joan Benavent, Xaro Benavent, Esther de Ves, Ruben Granados, and Ana García-Serrano 31

Multimedia Search with Noisy Modalities: Fusion and Multistage Retrieval
Avi Arampatzis, Savvas A. Chatzichristofis, and Konstantinos Zagoris 32

I2R At ImageCLEF Wikipedia Retrieval 2010
Kong-Wah Wan, Yan-Tao Zheng, and Sujoy Roy 33

DAEDALUS at ImageCLEF Wikipedia Retrieval 2010: Expanding with Semantic Information from Context
Sara Lana-Serrano, Julio Villena-Román, and José Carlos González-Cristóbal 34

SZTAKI@Imageclef2010
Bálint Daróczy, István Petrás, András A. Benczúr, Dávid Márk Nemeskey, and Róbert Pethes 34

RGU at ImageCLEF2010 Wikipedia Retrieval Task
Jun Wang, Dawei Song, and Leszek Kaliciak 35

UNT at ImageCLEF 2010: CLIR for Wikipedia Images
Miguel E. Ruiz, Jiangping Chen, Karthikeyan Pasupathy, Pok Chin, and Ryan Knudson 35

ImageCLEF: Medical Retrieval

ImageCLEF 2010 Working Notes on the Modality Classification Subtask
Olivier Pauly, Diana Mateus, and Nassir Navab 36

ImageCLEF 2010 Modality Classification in Medical Image Retrieval: Multiple Feature Fusion with Normalized Kernel Function
Xian-Hua Hanand Yen-Wei Chen 37

SINAI at ImageCLEF 2010 Medical Task
Manuel Carlos Díaz-Galiano, Maria Teresa Martín-Valdivia, Arturo Montejo-Ráez, and Miguel Angel García-Cumbreras 37

IRIT at ImageCLEF 2010: Medical Retrieval Track
Duy Dinh and Lynda Tamine 38

ImageCLEF 2010 Modality Classification in Medical Image Retrieval: Multiple Feature Fusion with Normalized Kernel Function

Xian-Hua Hanand Yen-Wei Chen

College of Information Science and Engineering, Ritsumeikan University, Kasatsu-shi, 525-8577, Japan
{hanxhua,chen}@{fc,is}.ritsumei.ac.jp

Our works are mainly to do modality classification for medical image retrieval task. The main objectives of the done experiments are to recognize what kind of modalities such as CT, MR, US and so on one input image possibly belongs to, which will be able to contribute to medical image retrieval. In our works, we extracted different visual features such as conventional gray/colour intensity histogram, variance and edge histogram and also the recently popular Bag-Of-Feature model for image representation, which are proved to be effective for general image categorization in computer vision. Furthermore, a binary histogram of key-words occurrence using images' captions as textual feature is explored for modality classification. After obtain the different features for image representation, we combine them together using kernel function in SVM classifier. Because different features maybe have deferent scale and dimension, in order to allow each individual feature to contribute equally for modality classification, we normalize the distance between two samples using mean distance of all training samples, and then, obtain the kernel function for each individual feature. The final kernel for SVM classification is the mean of individual kernel, which can be called Joint Kernel Equal Contribution (JKEC).
Our experiments use the database released for the ImageCLEF-2010 Medical modality classification in medical retrieval task, which includes 2390 annotated modality images (CT: 314; GX: 355; MR: 299; NM: 204; PET: 285; PX: 330; US: 307; XR:296) for training and a separate evaluated set consisting of 2620 images. The aim is to automatically classify the evaluated set using 8 different modality label sets including CT, MR, PET and so on. We used SVM classifier with visual, textual and combination of visual and textual features. The combined RBF kernel is calculated using the proposed Joint Kernel Equal Contribution (JKEC) for evaluating test image set. The recognition rates on test image set with visual, textual and combination of visual and textual features are 87.07%, 84.58% and 93.36%, respectively.

SINAI at ImageCLEF 2010 Medical Task

Manuel Carlos Díaz-Galiano, Maria Teresa Martín-Valdivia, Arturo Montejo-Ráez, and Miguel Angel García-Cumbreras

University of Jaén, Departamento de Informática, Jaén, Spain
{mcdiaz, maite, amontejo, magc}@ujaen.es

Recent researches demonstrate that the use and integration of several knowledge sources improves the quality and efficiency of information systems. In this paper, we present the system developed for the ImageCLEF 2010 medical task. We show the effect of using the medical ontology MeSH to expand terms found in textual queries.
This year we have applied a strategy for deciding when include the information extracted from MeSH. Two types of terms from the MeSH descriptors: MeSH Heading (MH), which is a term composed by one or more words. Each record contains only one MH term.
Entry, which is composed of one or more words too. Each record contains several Entry terms. These terms are a different way of writing the MH term, that is, they are synonyms.
In order to reduce the number of terms available to expand the query, we have only used those that are in the MeSH categories [A] Anatomy, [C] Diseases, and [E] Analytical, Diagnostic and Therapeutic Techniques and Equipment. The well known learning algorithms Support Vector Machine (SVM) was applied in these experiments. The learning approach consists, therefore, in training a binary classifier with part of the data (the queries of previous years) and test its performance with the queries tested in 2010.
The experiments carried out show that our machine learning approach to determine when to perform expansion did not resulted in any improvement over our base line, that where no expansion was performed at all.It is clear that our expansion strategy should be studied in depth, as the behaviour of the expansion over queries is not consistent and seems not related to the gain in MAP.