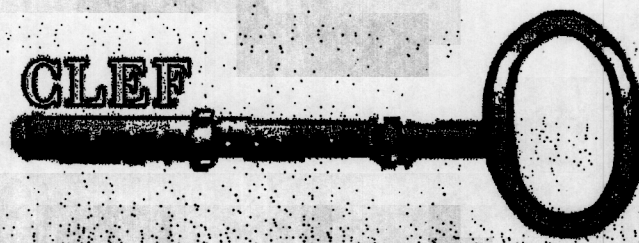


Cross Language Evaluation Forum



**Results of the CLEF 2005
Cross-Language System
Evaluation Campaign**

EXTENDED ABSTRACTS

CLEF 2005 Workshop, 21-23 September, Vienna, Austria

CONTENTS

What happened in CLEF 2005? <i>Carol Peters</i>	1
AD HOC: CROSS-LANGUAGE AND MONOLINGUAL	3
CLEF 2005: Ad Hoc Track Overview <i>Giorgio M. Di Nunzio, Nicola Ferro, Gareth J.F. Jones, Carol Peters</i>	3
Ad-hoc Mono- and Bilingual Retrieval Experiments at the University of Hildesheim. <i>René Hackl, Thomas Mandl, Christa Womser-Hacker</i>	4
MIRACLE's 2005 Approach to Cross-lingual Information Retrieval <i>José C. González, José Miguel Goñi-Menoyo, Julio Villena-Román</i>	5
The XLDB Group's Participation in the CLEF 2005 Ad Hoc Task <i>Nuno Cardoso, Leonardo Andrade, Alberto Simões, Mário J. Silva</i>	6
Thomson Legal and Regulatory Experiments at CLEF-2005 <i>Isabelle Moulinier, Ken Williams</i>	7
Using Syntactic Dependency and Language Model X-IOTA IR System for CLIPS Mono & Bilingual Experiments in CLEF 2005 <i>Loïc Maisonnasse, Gilles Sérasset, Jean-Pierre Chevallet</i>	8
Bilingual and Multilingual Experiments with the IR-n system <i>Elisa Noguera, Fernando Llopis, Rafael Muñoz, Rafael M. Terol, Miguel A. García-Cumbreras, Fernando Martínez-Santiago, Arturo Montejo-Raez</i>	9
Dictionary-based Amharic-French Information Retrieval <i>Atelach Alemu Argaw, Lars Asker, Rickard Cöster, Jussi Karlgren, Magnus Sahlgren</i>	10
Hybrid Approach to Query and Document Translation with Pivot Language for Cross-Language Information Retrieval <i>Kazuaki Kishida, Noriko Kando</i>	11
Ontology-Based Multilingual Information Retrieval <i>Jacques Guyot, Saïd Radhouani, Gilles Falquet</i>	12
SINAI at CLEF 2005: Multi-8 Two-years-on and Multi-8 Merging-only Tasks <i>Fernando Martínez-Santiago, Miguel A. García-Cumbreras</i>	13
CLEF 2005: Multilingual Retrieval by Combining Multiple Multilingual Ranked Lists <i>Luo Si, Jamie Callan</i>	14
Report on CLEF-2005 Evaluation Campaign: Monolingual, Bilingual, and GIRT Information Retrieval <i>Jacques Savoy, Pierre-Yves Berger</i>	15
Sociopolitical Thesaurus in Concept-based Information Retrieval <i>M. Ageev, B. Dobrov, N. Loukachevitch</i>	16
University of Indonesia's Participation in Ad Hoc at CLEF 2005 <i>Mirna Adriani, Ihsan Wahyu</i>	17
Exploring New Languages with HAIRCUT at CLEF 2005 <i>Paul McNamee</i>	18
Dublin City University at CLEF 2005: Multilingual Merging Experiments <i>Adenike M. Lam-Adesina, Gareth J. F. Jones</i>	19
Hungarian Monolingual Retrieval at CLEF 2005 <i>Anna Tordai, Maarten de Rijke</i>	20
ENSM-SE at CLEF 2005: Uses of Fuzzy Proximity Matching Function <i>Annabelle Mercier, Amelie Imafouo, Michel Beigbeder</i>	21
European Ad Hoc Retrieval Experiments with Hummingbird SearchServerTM at CLEF 2005 <i>Stephen Tomlinson</i>	22
Combining Passages in the Monolingual Task with the IR-n System <i>Fernando Llopis, Elisa Noguera</i>	23
Principled Query Processing <i>Jussi Karlgren, Magnus Sahlgren, Rickard Cöster</i>	24
MIRACLE's 2005 Approach to Monolingual Information Retrieval <i>José Miguel Goñi-Menoyo, José C. González, Julio Villena-Román</i>	25
Portuguese at CLEF 2005: Reflections and Challenges <i>Diana Santos, Nuno Cardoso</i>	26

DOMAIN-SPECIFIC INFORMATION RETRIEVAL.....	27
Fusion of Probabilistic Algorithms for the CLEF Domain Specific Task <i>Ray R. Larson</i>	27
Domain-Specific Russian Retrieval: A Baseline Approach <i>Fredric C. Gey</i>	28
University of Hagen at CLEF 2005: Towards a Better Baseline for NLP Methods in Domain-Specific Information Retrieval <i>Johannes Leveling</i>	29
How One Word Can Make all the Difference – Using Subject Metadata for Automatic Query Expansion and Reformulation <i>Vivien Petras</i>	30
Evaluating a Conceptual Indexing Method by Utilizing WordNet <i>Mustapha Baziz, Mohand Boughanem, Nathalie Aussenac-Gilles</i>	31
Mono- and Bilingual Retrieval Experiments with a Social Science Document Corpus <i>René Hackl, Thomas Mandl</i>	32
INTERACTIVE CROSS-LANGUAGE INFORMATION RETRIEVAL.....	33
The CLEF 2005 Interactive Track: Cross-Language Question Answering and Image Retrieval <i>Julio Gonzalo, Paul Clough, Alessandro Vallin</i>	33
iCLEF2005 at REINA-USAL: Use of Free On-line Machine Translation Programs for Interactive Cross-Language Question Answering <i>Ángel F. Zazo Rodríguez, Carlos G. Figuerola, José Luis Alonso Berrocal, Viviana Fernández Marcial</i>	34
“How much context do you need?” An Experiment about Context Size in Interactive Cross-language Question Answering <i>Borja Navarro, Lorenza Moreno-Monteagudo, Elisa Noguera, Sonia Vázquez, Fernando Llopis, Andrés Montoyo</i>	35
UNED at iCLEF 2005: Automatic Highlighting of Potential Answers <i>Víctor Peinado, Fernando López-Ostenero, Julio Gonzalo, Felisa Verdejo</i>	36
Boolean Operators in Interactive Search <i>Julio Villena-Román, Raquel M. Crespo-García, José Carlos González-Cristóbal</i>	37
Concept Hierarchy across Languages in Text-Based Image Retrieval: A User Evaluation <i>Daniela Petrelli, Paul Clough</i>	38
MULTIPLE LANGUAGE QUESTION ANSWERING.....	39
Overview of the CLEF 2005 Multilingual Question Answering Track <i>Alessandro Vallin, Danilo Giampiccolo, Lili Aunimo, Christelle Ayache, Petya Osenova, Anselmo Peñas, Maarten de Rijke, Bogdan Sacaleanu, Diana Santos, Richard Sutcliffe</i>	39
A Fast Forward Approach to Cross-lingual Question Answering for English and German <i>Robert Strötgen, Thomas Mandl, René Schneider</i>	40
The Oedipe System at CLEF-QA 2005 <i>Romarc Besançon, Mehdi Embarek, Olivier Ferret</i>	41
Building an XML Framework for Question Answering <i>David Tomás, José L. Vicedo, Maximiliano Saiz, Rubén Izquierdo</i>	42
A Logic Programming-based Approach to the QA@CLEF05 track <i>Paulo Quaresm, Irene Rodrigues</i>	43
University of Hagen at QA@CLEF 2005: Extending Knowledge and Deepening Linguistic Processing for Question Answering <i>Sven Hartrumpf</i>	44
Question Answering for Dutch using Dependency Relations <i>Gosse Bouma, Jori Mur, Gertjan van Noord, Lonneke van der Plas, Jörg Tiedemann</i>	45
Term Translation Validation by Retrieving Bi-terms <i>Brigitte Grau, Anne-Laure Ligozat, Isabelle Robba, Madeleine Sialeu, Anne Vilnat</i>	46
Exploiting Linguistic Indices and Syntactic Structures for Multilingual Question Answering: ITC-irst at CLEF 2005 <i>Hristo Tanev, Milen Kouylekov, Bernardo Magnini, Matteo Negri, Kiril Simov</i>	47
The TALP-QA System for Spanish at CLEF-2005 <i>Daniel Ferrés, Samir Kanaan, Edgar González, Alicia Ageno, Horacio Rodríguez, Jordi Turmo</i>	48

Priberam's Question Answering System for Portuguese <i>Carlos Amaral, Helena Figueira, André Martins, Afonso Mendes, Pedro Mendes, Cláudia Pinto</i>	49
INAOE-UPV Joint Participation in CLEF 2005: Experiments in Monolingual Question Answering <i>M. Montes-y-Gómez, L. Villaseñor-Pineda, M. Pérez-Coutiño, J. M. Gómez-Soriano, E. Sanchis-Arnal, P. Rosso</i>	50
DFKI's LT-lab at the CLEF 2005 Multiple Language Question Answering Track <i>Günter Neuman, Bogdan Sacaleanu</i>	51
Monolingual and Cross-language QA using a QA-oriented Passage Retrieval System <i>José Manuel Gómez Soriano, Empar Bisbal Asensi, Davide Buscaldi, Paolo Rosso, Emilio Sanchis Arnal</i>	52
The University of Amsterdam at QA@CLEF 2005 <i>David Ahn, Valentin Jijkoun, Karin Müller, Maarten de Rijke, Erik Tjong, Kim Sang</i>	53
AliQAn, Spanish QA System at CLEF-2005 <i>S. Roger, S. Ferrández, A. Ferrández, J. Peral, F. Llopis, A. Aguilar, D. Tomás</i>	54
20th Century Esfinge (Sphinx) Solving the Riddles at CLEF 2005 <i>Luis Costa</i>	55
Question Answering using Semantic Annotation <i>Lili Aunimo, Reeta Kuuskoski</i>	56
MIRACLE's 2005 Approach to Cross-Lingual Question Answering <i>César de Pablo-Sánchez, Ana González-Ledesma, José Luis Martínez-Fernández, José Maria Guirao, Paloma Martínez, Antonio Moreno</i>	57
Cross Lingual Question Answering using QRISTAL for CLEF 2005 <i>Dominique Laurent, Patrick Séguéla, Sophie Nègre</i>	58
Experiments for Tuning the Values of Lexical Features in Question Answering for Spanish <i>Manuel Pérez-Coutiño, Manuel Montes-y-Gómez, Aurelio López-López, Luis Villaseñor-Pineda</i>	59
Cross-Language French-English Question Answering using the DLT System at CLEF 2005 <i>Richard F. E. Sutcliffe, Michael Mulcahy, Igal Gabbay, Aoife O'Gorman, Kieran White, Darina Slattery</i>	60
University of Indonesia's Participation in Question Answering at CLEF 2005 <i>Mirna Adriani and R. Rinawati</i>	61
BulQA: Bulgarian--Bulgarian Question Answering at CLEF 2005 <i>Kiril Simov, Petya Osenova</i>	62
CROSS-LANGUAGE RETRIEVAL IN IMAGE COLLECTIONS	63
The CLEF 2005 Cross-Language Image Retrieval Track <i>Paul Clough, Henning Müller, Thomas Deselaers, Michael Grubinger, Thomas Lehmann, Jeffery Jensen, William Hersh</i>	63
Towards a Topic Complexity Measure for Cross Language Image Retrieval <i>Michael Grubinger, Clement Leung, and Paul Clough</i>	64
Dublin City University at CLEF 2005: Experiments with the ImageCLEF St Andrew's Collection <i>Gareth J. F. Jones, Kieran McDonald</i>	65
Exploiting Semantic Features for Image Retrieval at CLEF 2005 <i>J.L. Martínez-Fernández, J. Villena, Ana García-Serrano, S. González-Tortosa, F. Carbone, M. Castagnone</i>	66
UNED at ImageCLEF 2005: Automatically Structured Queries with Named Entities over Metadata <i>Víctor Peinado, Fernando López-Ostenero, Julio Gonzalo</i>	67
Recovering Translation Errors in Cross-Language Image Retrieval using Word Association Models <i>Masashi Inoue</i>	68
Combining Text and Image Queries at ImageCLEF2005 <i>Yih-Chen Chang, Wen-Cheng Lin, Hsin-Hsi Chen</i>	69
CUHK Experiments with ImageCLEF 2005 <i>Steven C.H. Hoi, Jianke Zhu, Michael R. Lyu</i>	70
SINAI at ImageCLEF 2005 <i>M.T. Martín-Valdivia, M.A. García-Cumbreras, M.C. Díaz-Galiano, L.A. Ureña-López, A. Montejo-Raez</i>	71
Merging Results from Different Media: Lic2m Experiments at ImageCLEF 2005 <i>Romario Besançon, Christophe Millet</i>	72
Combining Multilevel Visual Features for Medical Image Retrieval in ImageCLEFmed 2005 <i>Wei Xiong, Bo Qiu, Qi Tian, Changsheng Xu, S.H. Ong, Kelvin Foong</i>	73

A Structured Learning Approach for Medical Image Indexing and Retrieval <i>Joo-Hwee Lim, Jean-Pierre Chevallet</i>	74
FIRE in ImageCLEF 2005: Combining Content-based Image Retrieval with Textual Information Retrieval <i>Thomas Deselaers, Tobias Weyand, Daniel Keysers, Wolfgang Macherey, Hermann Ney</i>	75
Categorizing and Annotating Medical Images by Retrieving Terms Relevant to Visual Features <i>Desislava Petkova, Lisa Ballesteros</i>	76
Manual Query Modification and Automated Translation to Improve Cross-Language Medical Image Retrieval <i>Jeffery R. Jensen, William R. Hersh</i>	77
MIRACLE's Combination of Visual and Textual Queries for Medical Image Retrieval <i>Julio Villena-Román, José Carlos González-Cristóbal, José Miguel Goñi-Menoyo, José Luis Martínez-Fernandez, Juan José Fernández</i>	78
Supervised Machine Learning based Medical Image Annotation and Retrieval <i>Md. Mahmudur Rahman, Bipin C. Desai, Prabir Bhattacharya</i>	79
Combining Global features for Content-based Retrieval of Medical Images <i>Mark O Güld, Christian Thies, Benedikt Fischer, Thomas M Lehmann</i>	80
NCTU_DBLAB@ImageCLEFmed 2005: Medical Image Retrieval Task <i>Pei-Cheng Cheng, Been-Chian Chien, Hao-Ren Ke, Wei-Pang Yang</i>	81
Using medGIFT and easyIR for the ImageCLEF 2005 Evaluation Tasks <i>Henning Müller, Antoine Geissbühler, Johan Marty, Christian Lovis, Patrick Ruch</i>	82
The University of Indonesia's Participation in IMAGE-CLEF 2005 <i>Mirna Adriani, A. Framadhan</i>	83
MIRACLE's Naive Approach to Medical Images Annotation <i>Julio Villena-Román, José Carlos González-Cristóbal, José Miguel Goñi-Menoyo, José Luis Martínez-Fernandez</i>	84
Report on the Annotation Task in ImageCLEFmed 2005 <i>Bo Qiu, Wei Xiong, Qi Tian, Chang Sheng Xu</i>	85
Using Ontology Dimensions and Negative Expansion to solve Precise Queries in the ImageCLEF Medical Task <i>Jean-Pierre Chevallet, Joo-Hwee Lim, Saïd Radhouani</i>	86
UB at CLEF 2005: Medical Image Retrieval Task <i>Miguel E. Ruiz, Silvia Southwick</i>	87
NCTU_DBLAB@ImageCLEF 2005: Automatic Annotation Task <i>Pei-Cheng Cheng, Been-Chian Chien, Hao-Ren Ke, Wei-Pang Yang</i>	88
CROSS-LANGUAGE SPEECH RETRIEVAL	89
CLEF 2005 Cross-Language Speech Retrieval Track Overview <i>Ryen W. White, Douglas W. Oard, Gareth J. F. Jones, Dagobert Soergel, Xiaoli Huang</i>	89
University of Ottawa's Contribution to CLEF 2005, the CL-SR Track <i>Diana Inkpen, Muath Alzghool, Aminul Islam</i>	90
Waterloo Experiments for the CLEF05 SDR Track <i>Charles L. A. Clarke</i>	91
The University of Alicante at CL-SR track <i>Rafael M. Terol, Manuel Palomar, Patricio Martinez-Barco, Fernando Llopis, Rafael Muñoz, Elisa Noguera</i>	92
Pitt at CLEF05: Data Fusion for Spoken Document Retrieval <i>Daqing He, Jaewook Ahn</i>	93
UNED at CL-SR CLEF 2005: Mixing Different Strategies to Retrieve Automatic Speech Transcriptions <i>Fernando López-Ostenero, Víctor Peinado, Valentín Sama</i>	94
Dublin City University at CLEF 2005: Cross-Language Spoken Document Retrieval (CL-SR) Experiments <i>Adenike M. Lam-Adesina, Gareth J. F. Jones</i>	95
CLEF-2005 CL-SR at Maryland: Document and Query Expansion using Side Collections and Thesauri <i>Jianqiang Wang, Douglas W. Oard</i>	96

MULTILINGUAL WEB TRACK.....	97
Overview of WebCLEF 2005	
<i>Börkur Sigurbjörnsson, Jaap Kamps, Maarten de Rijke.....</i>	<i>97</i>
EuroGOV: Engineering a Multilingual Web Corpus	
<i>Börkur Sigurbjörnsson, Jaap Kamps, Maarten de Rijke.....</i>	<i>98</i>
Web Retrieval Experiments with the EuroGOV Corpus at the University of Hildesheim	
<i>Niels Jensen, René Hackl, Thomas Mandl, Robert Strötgen.....</i>	<i>99</i>
European Web Retrieval Experiments with Hummingbird SearchServer™ at CLEF 2005	
<i>Stephen Tomlinson.....</i>	<i>100</i>
Melange: Components for Cross-Lingual Retrieval	
<i>Max Pfingsthorn, Koen van de Sande, Vladimir Nedovic.....</i>	<i>101</i>
The University of Amsterdam at WebCLEF 2005	
<i>Jaap Kamps, Maarten de Rijke, Borkur Sigurbjörnsson.....</i>	<i>102</i>
Web Track for CLEF2005 at ALICANTE UNIVERSITY	
<i>Trinitario Martínez, Elisa Noguera, Rafael Muñoz, Fernando Llopis.....</i>	<i>103</i>
MIRACLE's Approach to Multilingual Web Retrieval	
<i>Ángel Martínez-González, José Luis Martínez-Fernández, César de Pablo-Sánchez, Julio Villena-Román, Luis Jiménez-Cuadrado, Paloma Martínez, José Carlos González-Cristóbal.....</i>	<i>104</i>
TPIRS: A System for Document Indexing Reduction on WebCLEF	
<i>David Pinto, Héctor Jiménez-Salazar, Paolo Rosso, Emilio Sanchis.....</i>	<i>105</i>
REINA at the WebCLEF Task: Combining Evidences and Link Analysis	
<i>Carlos G. Figuerola, José L. Alonso Berrocal, Ángel F. Zazo Rodríguez, Emilio Rodríguez.....</i>	<i>106</i>
UNED at WebCLEF 2005	
<i>Javier Artiles, Víctor Peinado, Anselmo Penas, Felisa Verdejo.....</i>	<i>107</i>
University of Indonesia's Participation in WEB-CLEF 2005	
<i>Mirna Adriani, Rama Pandugita.....</i>	<i>108</i>
CROSS-LANGUAGE GEOGRAPHICAL RETRIEVAL.....	109
GeoCLEF: the CLEF 2005 Cross-Language Geographic Information Retrieval Track	
<i>Fredric Gey, Ray Larson, Mark Sanderson, Hideo Joho, Paul Clough.....</i>	<i>109</i>
MIRACLE's 2005 Approach to Geographical Information Retrieval	
<i>Sara Lana-Serrano, José M. Goñi-Menoyo, José C. González-Cristóbal.....</i>	<i>110</i>
University of Alicante at GeoCLEF 2005	
<i>O. Ferrández, Z. Kozareva, A. Toral, E. Noguera, A. Montoyo, R. Muñoz Fernando Llopis.....</i>	<i>111</i>
MetaCarta at GeoCLEF 2005	
<i>András Kornai.....</i>	<i>112</i>
A WordNet-based Query Expansion Method for Geographical Information Retrieval	
<i>Davide Buscaldi, Paolo Rosso, Emilio Sanchis Arnal.....</i>	<i>113</i>
The GeoTALP-IR System at GeoCLEF-2005: Experiments Using a QA-based IR System, Linguistic Analysis, and a Geographical Thesaurus	
<i>Daniel Ferrés, Alicia Ageno, Horacio Rodríguez.....</i>	<i>114</i>
Cheshire II at GeoCLEF: Fusion and Query Expansion for GIR	
<i>Ray R. Larson.....</i>	<i>115</i>
CSUSM Experiments in GeoCLEF2005: Monolingual and Bilingual Tasks	
<i>Rocio Guillén.....</i>	<i>116</i>
Berkeley2 at GeoCLEF: Cross-Language Geographic Information Retrieval of German and English Documents	
<i>Vivien Petras, Fredric Gey.....</i>	<i>117</i>
University of Hagen at GeoCLEF 2005: Using Semantic Networks for Interpreting Geographical Queries	
<i>Johannes Leveling, Sven Hartrumpf, Dirk Veiel.....</i>	<i>118</i>
Preliminary Experiments with Geo-Filtering Predicates for Geographic IR	
<i>Jochen L. Leidner.....</i>	<i>119</i>
NICTA i2d2 in GeoCLEF 2005	
<i>Baden Hughes.....</i>	<i>120</i>
PARTICIPATING INSTITUTIONS.....	121
CLEF STEERING COMMITTEE.....	123
ACKNOWLEDGMENTS.....	124

Bilingual and Multilingual Experiments with the IR-n System

Elisa Noguera¹, Fernando Llopis¹, Rafael Muñoz¹,
Rafael M. Terol¹, Miguel A. García-Cumbreras², Fernando Martínez-Santiago²,
and Arturo Montejo-Raez²

¹ Grupo de investigación en Procesamiento del Lenguaje Natural y Sistemas de
Información

Departamento de Lenguajes y Sistemas Informáticos
University of Alicante, Spain

{`elisa`, `llopis`, `rafael`, `rafaelmt`}@`dlssi.ua.es`

² Department of Computer Science. University of Jaen, Jaen, Spain
{`magc`, `dofer`, `montejo`}@`ujaen.es`

Abstract. Our paper describes the participation of the IR-n system at CLEF-2005. This year, we participated in the bilingual task (English-French and English-Portuguese) and the multilingual task (English, French, Italian, German, Dutch, Finish and Swedish). We introduced the method of combined passages for the bilingual task. Furthermore we have applied the method of logic forms in the same task. For the multilingual task we had a joint participation with the University of Alicante and University of Jaén. We want to emphasize the good score achieved in the bilingual task improving around 45% in terms of average precision.

1 Introduction

Information Retrieval (IR) systems [2] try to find the relevant documents given a user query from a document collection. We can find different types of IR systems in the literature. On the one hand, if the document collection and the user query are written in the same language then the IR system can be defined as a monolingual IR system. On the other hand, if the document collection and the user query are written in different languages then the IR system can be defined as a bilingual (two different languages) or multilingual (more than two languages) IR system. Obviously, the document collection for multilingual systems is written in at least two different languages. The IR-n system [3] can work with collections and queries in any language.

Passage Retrieval (PR) systems [1] are information retrieval systems that determine the similarity of a document with regard to a user query according to the similarity of fragments of the document (passages) with regard to the same query.

2 Bilingual Task

2.1 Method 1: Machine Translation

We use different translators in order to obtain an automatic translation of queries. Three of them were used for all languages: FreeTranslation, Babel Fish and InterTran. Moreover, we have used one more method merging all translations. This is performed by merging several translations from the on-line translators. This strategy is based on the idea that the words which appear in multiple translations have more relevancy than those that only appear in one translation.

The method of combined passages was developed in the monolingual task [4], for this reason it has been also used in the bilingual task. In the training test, the best input configuration has been used for French and Portuguese. Best scores were achieved using the merge of translations in English-Portuguese and FreeTranslation in English-French.

2.2 Method 2: Logic Forms

The last release of our IR-n system introduced a set of features that are based on the application of logic forms to topics and in the incrementing of the weight of terms according to a set of syntactic rules. The reason for this is that IR-n system includes a new module that increments the weight of terms, applying a set of rules based on the representation of the topics in the way of logic forms [7].

3 Multilingual Task

This year we made a combination between the fusion algorithm 2-step RSV [6], developed by the University of Jaén, and the passage retrieval system IR-n, developed by the University of Alicante. A full detailed description of the experiments is available in this volume.

IR-n has been used as IR system in order to make some experiments in Multi-8 Two-years-on task. Thus, it has been applied over eight languages: English, Spanish, French, Italian, German, Dutch, Finnish and Swedish.

An in depth description of the training test is available in [6]. Firstly, each monolingual collection is preprocessed as usual (token extraction, stopwords are eliminated and stemming is applied to the rest of words). In addition, compound words are decompounded as possible for German, Swedish, Finnish and Dutch. We use the decompounding algorithm depicted in [5]. The preprocessed collections have been indexed using the passage retrieval system IR-n and the document retrieval system ZPrise. The IR-n system has been modified in order to return a list of the retrieved and relevant documents, the documents that contain the relevant passages. Finally, given a query and its translations into the other languages, each query is searched in the corresponding monolingual collection.

When the monolingual lists of relevant documents are returned, we apply the 2-step RSV fusion algorithm. This algorithm deals with the translations

whose terms are known (aligned terms) in a different way than those words whose translation is unknown (non-aligned words) by giving two scores for each document. The first one is calculated taking into account aligned words, and the second one only uses non-aligned terms. Thus, both scores are combined into a only RSV per document and query by using some formulae:

1. Combining the RSV value of the aligned words and not aligned words with the formula:

$$0.6 < RSV AlignedDoc > + 0.4 < RSV NotAligned >$$

2. By using Logistic Regression. The formula:

$$e^{\alpha \cdot < RSV AlignedDoc > + \beta \cdot < RSV NotAligned >}$$

3. The last one also uses Logistic Regression but include a new component, the ranking of the doc. It applies the formula:

$$e^{\alpha \cdot < RSV AlignedDoc > + \beta \cdot < RSV NotAligned > + \gamma \cdot < RankingDoc >}$$

4 Results at CLEF-2005

The IR-n system used the best configuration obtained in the training process. Three different runs have been submitted for each task. The first run IRn-xx-vep uses the method of combined passages with query expansion. The second run IRn-xx-fexp only uses query expansion. The third run IRn-xx-vnexp uses the method of combined passages without query expansion. Furthermore, a fourth run IRn-xx-fexpfl has been submitted for English-Portuguese task. It uses the method of logic forms. Table 1 shows the scores achieved for each run.

Table 1 shows the official results for "Multi-8 Two-years on task. IR-n performs better than ZPrise except for Finnish results, the differences of average precision between both multilingual experiments is very small. The reason is that the merging algorithm is independent of the initial selection of relevant documents. This feature has been briefly discussed above and in more detail in [6].

5 Conclusions and Future Work

In the bilingual task the IR-n system has obtained better results merging translations than using single translations. On the other hand, the method of combined passages improves the scores in the bilingual task compared to the method of fixed passages, as it happens in the monolingual task.

Thus, in the multilingual task we conclude that IR-n is a good information retrieval system for CLIR systems. It improves on document-based systems such as OKAPI-ZPrise in bilingual experiments. In addition, the integration of this system with complex merging algorithms such as 2-step RSV is straightforward.

Possibly, if an IR-n like system were implemented for the creation of a dynamic index the multilingual results would be improved in the same way that the monolingual results are.

Table 1. CLEF 2005 official results. Bilingual and Multilingual tasks.

BILINGUAL TASK			
Language	Run	AvgP	Dif
English - Portuguese	CLEF Average	21.71	+34.4%
	IRn-enpt-vexp	29.18	
	IRn-enpt-fexp	28.94	
	IRn-enpt-vnexp	25.22	
	IRn-enpt-fexpfl	27.27	
English - French	CLEF Average	24.76	+45.3%
	IRn-fr-vexp	35.90	
	IRn-fr-fexp	29.12	
	IRn-fr-vnexp	29.13	

MULTILINGUAL TASK			
IR system	Formula 1	Formula 2	Formula 3
ZPrise+OKAPI	28.78	29.01	29.12
IR-n	28.85	29.09	29.18

Acknowledgements

This research has been partially funded by the Spanish Government under project CICYT number TIC2003-07158-C04-01 and by the Valencian Government under project numbers GV04B-276 and GV04B-268.

References

1. M. Kaskziel and J. Zobel: Passage Retrieval Revisited. In *Proceedings of the 20th annual ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 97)*, Philadelphia, pages 178–185, 1997.
2. F. W. Lancaster: *Information Retrieval Systems: Characteristics, Testing and Eval.*, John Wiley and Sons, New York, 1979.
3. F. Llopis: *IR-n: Un Sistema de Recuperación de Información Basado en pasajes*. PhD thesis. University of Alicante, 2003.
4. F. Llopis and E. Noguera: Combining Passages in Monolingual Experiments with IR-n System. In *Workshop of Cross-Language Evaluation Forum (CLEF 2005)*, Vienna, In this volume.
5. F. Martínez-Santiago, M. García-Cumbreras and L. A. Ureña: SINAI at CLEF 2004: Using Machine Translation Resources with Mixed 2-Step RSV Merging Algorithm. *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum*, Bath, page 156-164, 2005.
6. F. Martínez-Santiago, L. A. Ureña, and M. Martín: A Merging Strategy Proposal: Two Step Retrieval Status Value Method. *Information Retrieval*, 9(1):95–109, 2006.
7. R. M. Terol: The University of Alicante at CL-SR Track. In *Workshop of Cross-Language Evaluation Forum (CLEF 2005)*, Vienna, In this volume.