CriES Workshop @CLEF 2010

Cross-lingual Expert Search - Bridging CLIR and Social Media

Organizing Committee:
Philipp Sorg
Antje Schultz
Philipp Cimiano
Sergej Sizov
# Workshop Program

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
</table>
| 14:30 | Introduction and Overview of the CriES Pilot Challenge<br *

_Philipp Sorg, Karlsruhe Institute of Technology_ |

| 15:00 | HITS and Misses: Combining BM25 with HITS for Expert Search.<br _n

_Johannes Leveling and Gareth J. F. Jones_ |

| 15:30 | Identify Experts from a Domain of Interest.<br _n

_Adrian Iftene, Bogdan Luca, Georgiana Cărăuşu, and Madălina Merchez_ |

_Coffee Break_

| 16:30 | Multilingual Expert Search using Linked Open Data as Interlingual Representation.<br _n

_Daniel M. Herzig and Hristina Taneva_ |

| 17:00 | Expertise Retrieval: Tasks, Methods, Evaluation<br _n

_Krisztian Balog_ |

| 17:50 | Wrap up and Final Discussion                                                                 |

---
Overview of the CriES Pilot Challenge: Dataset, Topics and Results

CRIES PILOT CHALLENGE
Outline

- (Brief) Introduction to Multilingual Expert Search
- CriES Pilot Challenge
  - Dataset
  - Topics
  - Evaluation
- Baseline Retrieval Approaches
- Results of Participants
- Lessons Learned
Motivation

People want to ask other people to satisfy their information needs
- Instead of searching themselves
- Possibly more efficient for very specific information needs
- Includes social aspects
- Prominent examples are Question/Answer portals
  - Yahoo! Answers, WikiAnswers

Many Community Portals are multilingual
- Large networks of users from many countries
- Communication only depends on common language (e.g. English)
Multilingual Expert Search

- Special Case of Entity Search
  - Entities = People (Experts)

- CriES Context
  - Topic = Information need of user
  - Experts are able to answer information need
    - No retrieval of the actual answers
  - Multilingual evidence
    - Assumption: All experts are able to communicate with all users
    - Expertise independent from language
  - Social features
    - Dataset from community portal
CriES Pilot Challenge

Main Problem

- Multilingual expert search in social media environments

Key Research Challenges:

User characterization

- The use of multilingual evidence (including text) of social media for building expert profiles.

Community analysis

- Mining of social relationships in collaborative environments for multilingual retrieval scenarios.

User-centric recommender algorithms

- Development of retrieval and recommendation algorithms that allow for similarity search and ranked retrieval of expert users in online communities.
Related Challenges

- TREC
  - Enterprise Track 2005 - 2008
  - Entity Track 2009 – 2010
  - INEX

- What are the differences to CriES?
  - Multilingual Evidence
    - Profiles
    - Topics
  - Features from Social Community Portal
    - Relations between Users
    - (Ratings and Reputation)
  - No focus on property extraction
Yahoo! Answers crawl used for the CriES challenge, Topics and Relevance Assessments

DATASET
Yahoo! Answers

- Community Question/Answer Portal
  - Users post questions
  - Other users answer questions

- Several Answers per Question

- Selection of Best Answers
  - By the questioner
  - Based on answer ratings of other users

- Features of a Social Community Portal
  - Ratings
  - Reputation
  - Personal contacts
Example from Yahoo! Answers

Resolved Question

Why do we 'yawn'?

Best Answer - Chosen by Asker

New research suggests that it doesn't have to do with breath...
Dataset

- Yahoo! Research Webscope program
  - L6. Yahoo! Answers Comprehensive Questions and Answers (version 1.0)

Features

- Questions (with user ids)
- Best answers (with user ids)
- Other answers (without user ids)
- Categories of questions

Missing Features

- Ratings for questions/answers
- Most features from community portal
CriES Dataset

- Properties of the Yahoo! Answers Dataset
  - Many questions have purpose of diversion
  - Many questions ask for opinions
  - Problems for expert retrieval scenario
    - Noise in user profiles
    - Relevance assessment might not be objective

- Solution
  - Use subset with suitable properties
  - Selected subset
    - Technical Categories
    - Require domain expertise
    - Low share of questions with the purpose of diversion
Topic and Language Distribution

- Selection based on Categories
  - Computer & Internet, Health, Science & Mathematics
  - Questions in English, German, French and Spanish
Topics

- Real User Needs
  - Usage of questions from dataset

- Multilingual Search Task
  - Select topics in all relevant languages
  - 15 topics each in English, German, French and Spanish

- Topics suitable for Retrieval Scenario
  - Definition of topic criteria
  - Selection process using manual assessments
Topic Selection Process

Criteria for Topics

- International domain
  - Why doesn't an optical mouse work on a glass table?
  - Why is it so foggy in San Francisco?
- Expertise questions
  - What is a blog?
  - What is the best podcast to subscribe to?

Selection Process

- 100 random questions in each language
- Manual assessment in respect to criteria
- Check for language coverage in dataset
  - Average number of matching answers (on term level)
Relevance Assessment

- Result Pool of Submitted Runs
  - Top 10 experts for each topic

- Manual Assessment
  - Using text profiles of experts

- 3 Relevance Classes
  - Expert is likely able to answer. *(Strict evaluation)*
  - Expert may be able to answer. *(Lenient evaluation)*
  - Expert is probably not able to answer.

- Additional Automatic Assessments
  - Questioner (non relevant) and answerer (relevant) of the questions used as topics
Question: Is there a cure for malaria?
Statistics of Relevance Assessment

- **Assessors**
  - 6 assessors
  - Students at KIT
  - Evaluation of 7,515 pairs of topics and expert profiles

- **Distribution**
  - 1678 (relevant)
  - 1864 (probably relevant)
  - 3973 (non relevant)

- **Distribution over Languages**
  - Bias towards topic language
Relevant Expert Language Distribution

English Topics

German Topics

French Topics

Spanish Topics
Baselines and Submitted Runs

RETRIEVAL APPROACHES
Baseline Retrieval Approaches

- Multilingual IR
  - Language specific indexes
  - Text profiles of experts
    - Former answers in each language
  - Standard retrieval model and aggregation
    - BM25
    - Z-Score normalization

- Category Baseline
  - Informed approach
    - Category of question is known
  - Measure „importance“ of experts in categories
    - Number of answers
    - Pagerank based on questioner / answerer relation
  - Return experts ranked by importance in question category
Summary of Approaches

- **MLIR Approach**
  - Query translation
    - Using Machine Translation Systems (e.g. Google Translate)
  - IR models used for ranking
    - Vector Space Models
    - Probabilistic Models (e.g. BM25)

- **Matching Runs**
  - Adrian Iftene et al.
    - Run0 and run1
  - Johannes Leveling and Gareth J. F. Jones
    - BM25 with various translation strategies
Summary of Approaches (2)

- Approaches based on Social Features
  - Building of social graph from dataset
    - Questions and answers
    - Categories
    - Questioners and answerers
  - Scoring functions for nodes
    - HITS
    - Degree of nodes

- Matching Runs
  - Adrian Iftene et al.
    - Run2
  - Johannes Leveling and Gareth J. F. Jones
    - Combination of HITS with BM25 scoring model
Summary of Approaches (3)

Resource Indexing
- Inter-lingual concept space
  - Allows for multilingual retrieval
- Mapping of experts to concept space
  - Based on expert profiles
- Mapping of topics to concept space

Matching Runs
- Daniel M. Herzig and Hristina Taneva
  - Wikipedia as concept space
  - Different approaches to build expert profiles
  - Manual and automatic mapping of topics
## Results

<table>
<thead>
<tr>
<th>Run Id</th>
<th>Strict</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P@10</td>
<td>MRR</td>
<td>P@10</td>
<td>MRR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social Features</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iftene (run2)</td>
<td>.62</td>
<td>.84</td>
<td>.83</td>
<td>.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category Baseline</td>
<td>.67</td>
<td>.89</td>
<td>.79</td>
<td>.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multilingual IR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iftene (run0)</td>
<td>.52</td>
<td>.80</td>
<td>.82</td>
<td>.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bastings</td>
<td>.07</td>
<td>.15</td>
<td>.25</td>
<td>.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM25 + Z-Score</td>
<td>.19</td>
<td>.40</td>
<td>.39</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MLIR + Social Features</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leveling (DCUq)</td>
<td>.08</td>
<td>.16</td>
<td>.42</td>
<td>.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Resource Indexing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herzig (3-boe-07-02-01-q01m)</td>
<td>.49</td>
<td>.76</td>
<td>.87</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results of Participants (Strict)
Results of Participants (Lenient)

- bastings
- herzig_1-boe-06-03-01-q01m
- herzig_2-boe-06-03-01-q01
- herzig_3-boe-07-02-01-q01m
- iftene_run0
- iftene_run1
- iftene_run2
- leveling_DCUa
- leveling_DCUq
Lessons Learned and Outlook

SUMMARY
Lessons Learned

- Observations
  - Assessors may be biased to positively judge experts with many answers
    - High evaluation values of category baseline
- Possible explanation
  - Design of user interface for assessments
    - Complete expert profiles
    - Could be “overwhelming”
- Alternative
  - Compare single answers of experts to topics
  - Problem: Expertise in context of several answers
Lessons Learned (2)

- Social features only used by some participants
  - Maybe dataset doesn’t support this approach?

- More features needed
  - Ratings of answers
    - Identify helpful answers
    - Use alternative answers for retrieval
  - Social profiles of users
    - History of best answers
    - Status in portal
    - Explicit relations to other users
Outlook

Classifier

Social Features

Category
Baseline

Pilot Challenge Participants

Text Features

IR enhanced with social features

Combined Classifier

MLIR
Thank you for your attention!

Questions?

- Acknowledgements
  - Multipla Project
    - http://www.multipla-project.org
  - Monnet Project
    - http://www.monnet-project.eu/