

# CLEF 2009: Grid@CLEF Pilot Track Overview

Nicola Ferro<sup>1</sup> and Donna Harman<sup>2</sup>

<sup>1</sup> Department of Information Engineering, University of Padua, Italy  
`ferro@dei.unipd.it`

<sup>2</sup> National Institute of Standards and Technology (NIST), USA  
`donna.harman@nist.gov`

**Abstract.** The Grid@CLEF track is a long term activity with the aim of running a series of systematic experiments in order to improve the comprehension of MLIA systems and gain an exhaustive picture of their behaviour with respect to languages.

In particular, Grid@CLEF 2009 is a pilot track that has started to move the first steps in this direction by giving the participants the possibility of getting experienced with the new way of carrying out experimentation that is needed in Grid@CLEF to test all the different combinations of IR components and languages. Grid@CLEF 2009 offered traditional monolingual ad-hoc tasks in 5 different languages (Dutch, English, French, German, and Italian) which make use of consolidated and very well known collections from CLEF 2001 and 2002 and used a set of 84 topics.

Participants had to conduct experiments according to the CIRCO framework, an XML-based protocol which allows for a distributed, loosely-coupled, and asynchronous experimental evaluation of IR systems. We provided a Java library which can be exploited to implement CIRCO and an example implementation with the Lucene IR system.

The participation has been especially challenging also for the size of the XML files generated by CIRCO, which can become 50-60 times the size of the collection. Of the 9 initially subscribed participants, only 2 were able to submit runs in time and we received a total of 18 runs in 3 languages (English, French, and German) out of the 5 offered. The two participants used different IR systems or combination of them, namely Lucene, Terrier, and Cheshire II.

## 1 Introduction

Much of the effort of *Cross-Language Evaluation Forum (CLEF)* over the years has been devoted to the investigation of key questions such as “What is *Cross Language Information Retrieval (CLIR)*?”, “What areas should it cover?” and “What resources, tools and technologies are needed?” In this respect, the Ad Hoc track has always been considered as the core track in CLEF and it has been the starting point for many groups as they begin to be interested in developing functionality for the multilingual information access. Thanks to this pioneering work, CLEF produced, over the years, the necessary groundwork and foundations

to be able, today, to start wondering how to go deeper and to address even more challenging issues [11,12].

The Grid@CLEF Pilot track<sup>1</sup> moves the first steps in this direction and aims at [10]:

- looking at differences across a wide set of languages;
- identifying best practices for each language;
- helping other countries to develop their expertise in the *Information Retrieval (IR)* field and create IR groups;
- providing a repository, in which all the information and knowledge derived from the experiments undertaken can be managed and made available via the *Distributed Information Retrieval Evaluation Campaign Tool (DIRECT)* system.

The Grid@CLEF pilot track in CLEF 2009 has provided us with an opportunity to begin to set up a suitable framework in order to carry out a first set of experiments which allows us to acquire an initial set of measurements and to start to explore the interaction among IR components and languages. This initial knowledge will allow us to tune the overall protocol and framework, to understand what directions are more promising, and to scale the experiments up to a finer-grain comprehension of the behaviour of IR components across languages.

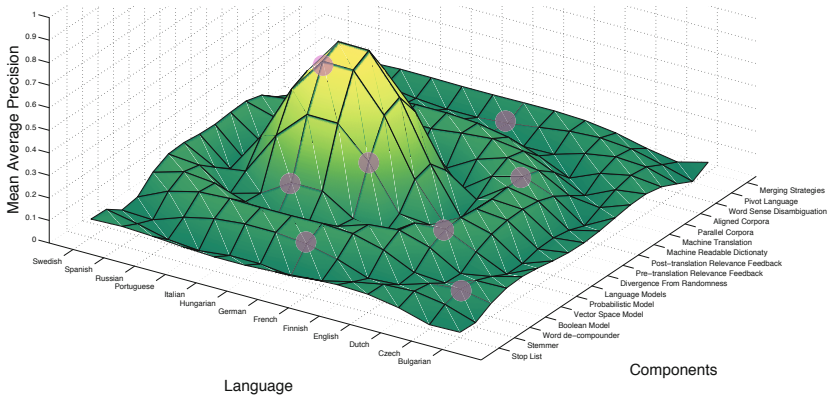
The paper is organized as follows: Section 2 provides an overview of the approach and the issues that need to be faced in Grid@CLEF; Section 3 introduces CIRCO, the framework we are developing in order to enable the Grid@CLEF experiments; Section 4 describes the experimental setup that has been adopted for Grid@CLEF 2009; Section 5 presents the main outcomes of this year Grid@CLEF in terms of participation and performances achieved; finally, Section 6 discusses the different approaches and findings of the participants in Grid@CLEF.

## 2 Grid@CLEF Approach

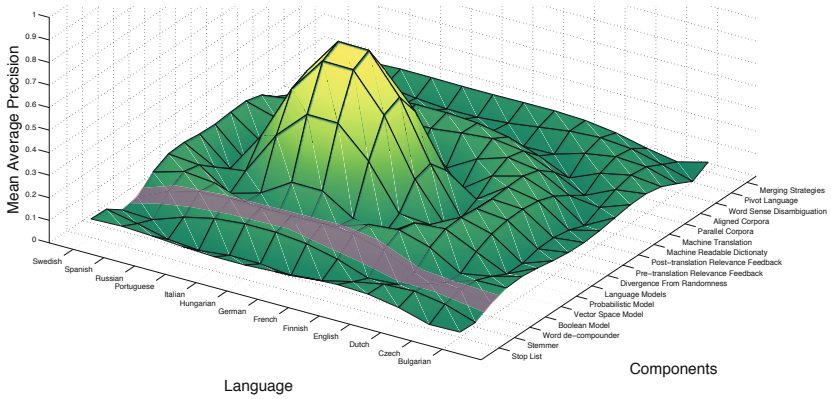
Individual researchers or small groups do not usually have the possibility of running large-scale and systematic experiments over a large set of experimental collections and resources. Figure 1 depicts the performances, e.g. mean average precision, of the composition of different IR components across a set of languages as a kind of surface area which we intend to explore with our experiment. The average CLEF participants, shown in Figure 1(a), may only be able to sample a few points on this surface since, for example, they usually test just a few variations of their own or customary IR model with a stemmer for two or three languages. Instead, the expert CLEF participant, represented in Figure 1(b), may have the expertise and competence to test all the possible variations of a given component across a set of languages, as [22] does for stemmers, thus investigating a good slice of the surface area.

---

<sup>1</sup> <http://ims.dei.unipd.it/gridclef/>

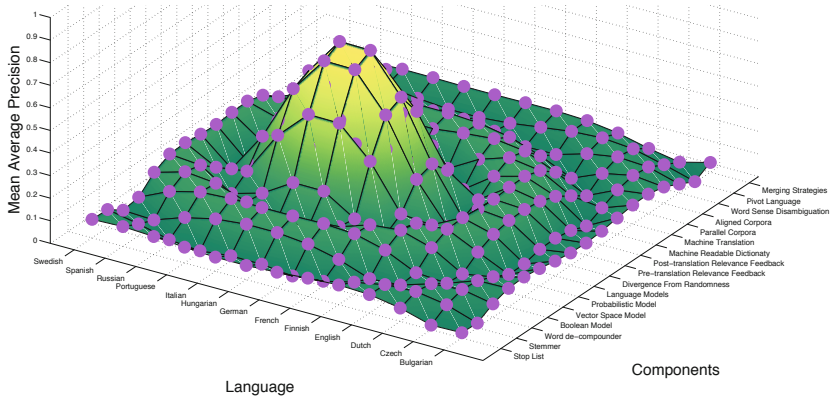


(a) Average CLEF participants.



(b) Expert CLEF participant.

**Fig. 1.** Coverage achieved by different kinds of participants



**Fig. 2.** The three main entities involved in grid experiments

However, even though each of these cases produces valuable research results and contributes to the advancement of the discipline, they are both still far removed from a clear and complete comprehension of the features and properties of the surface. A far deeper sampling would be needed for this, as shown in Figure 2: in this sense, Grid@CLEF will create a fine-grained *grid of points* over this surface and, hence, the name of the track comes.

It is our hypothesis that a series of systematic experiments can re-use and exploit the valuable resources and experimental collections made available by CLEF in order to gain more insights about the effectiveness of, for example, the various weighting schemes and retrieval techniques with respect to the languages.

In order to do this, we must deal with the interaction of three main entities:

- **Component:** in charge of carrying out one of the steps of the IR process;
- **Language:** will affect the performance and behaviour of the different components of an *Information Retrieval System (IRS)* depending on its specific features, e.g. alphabet, morphology, syntax, and so on.
- **Task:** will impact on the performances of IRS components according to its distinctive characteristics;

We assume that the contributions of these three main entities to retrieval performance tend to overlap; nevertheless, at present, we do not have enough knowledge about this process to say whether, how, and to what extent these entities interact and/or overlap – and how their contributions can be combined, e.g. in a linear fashion or according to some more complex relation.

The above issue is in direct relationship with another long-standing problem in the IR experimentation: the impossibility of testing a single component independently of a complete IRS. [16, p. 12] points out that “if we want to decide between alternative indexing strategies for example, we must use these strategies *as part of a complete information retrieval system*, and examine its overall performance (with each of the alternatives) directly”. This means that we have to proceed by changing only one component at time and keeping all the others fixed, in order to identify the impact of that component on retrieval effectiveness;

this also calls for the identification of suitable baselines with respect to which comparisons can be made.

### 3 The CIRCO Framework

In order to run these grid experiments, we need to set up a framework in which participants can exchange the intermediate output of the components of their systems and create a run by using the output of the components of other participants.

For example, if the expertise of participant A is in building stemmers and decompounders while participant B's expertise is in developing probabilistic IR models, we would like to make it possible for participant A to apply his stemmer to a document collection, pass the output to participant B, who tests his probabilistic IR model, thus obtaining a final run which represents the test of participant A's stemmer + participant B probabilistic IR model.

To this end, the objective of the *Coordinated Information Retrieval Components Orchestration (CIRCO)* framework [9] is to allow for a *distributed, loosely-coupled, and asynchronous* experimental evaluation of *Information Retrieval (IR)* systems where:

- *distributed* highlights that different stakeholders can take part to the experimentation each one providing one or more components of the whole IR system to be evaluated;
- *loosely-coupled* points out that minimal integration among the different components is required to carry out the experimentation;
- *asynchronous* underlines that no synchronization among the different components is required to carry out the experimentation.

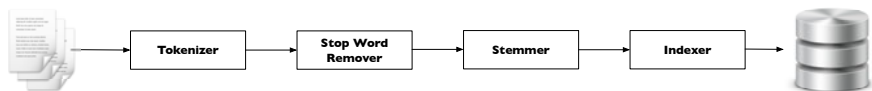
The CIRCO framework allows different research groups and industrial parties, each one with their own areas of expertise, to take part in the creation of *collaborative experiments*. This is a radical departure from today's IR evaluation practice where each stakeholder has to develop (or integrate components to build) an entire IR system to be able to run a single experiment.

The base idea – and assumption – behind CIRCO to streamline the architecture of an IR system and represent it as a *pipeline* of components chained together. The processing proceeds by passing the results of the computations of a component as input to the next component in the pipeline without branches, i.e. no alternative paths are allowed in the chain.

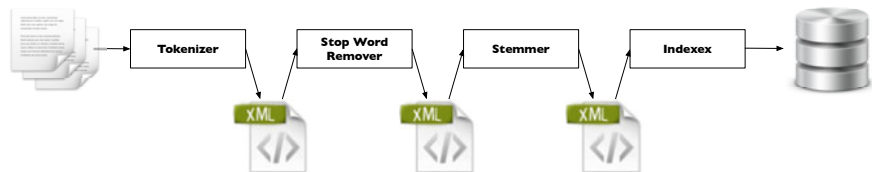
To get an intuitive idea of the overall approach adopted in CIRCO, consider the example pipeline shown in Figure 3(a).

The example IR system is constituted by the following components:

- *tokenizer*: breaks the input documents into a sequence of tokens;
- *stop word remover*: removes stop words from the sequence of tokens;
- *stemmer*: stems the tokens;
- *indexer*: weights the tokens and stores them and the related information in an index.



(a) An example pipeline for an IR system.



(b) An example of CIRCO pipeline for an IR system.

**Fig. 3.** Example of CIRCO approach to distributed, loosely-coupled, and asynchronous experimentation

Instead of directly feeding the next component as usually happens in an IR system, CIRCO operates by requiring each component to input and output from/to *eXtensible Markup Language (XML)* [25] files in a well-defined format, as shown in Figure 3(b).

These XML files can then be exchanged among the different stakeholders that are involved in the evaluation. In this way, we can meet the requirements stated above by allowing for an experimentation that is:

- *distributed* since different stakeholders can take part in the same experiment, each one providing his own component(s);
- *loosely-coupled* since the different components do not need to be integrated into a whole and running IR system but only need to communicate by means of a well-defined XML format;
- *asynchronous* since the different components do not need to operate all at the same time or immediately after the previous one but can exchange and process the XML files at different rates.

In order to allow this way of conducting experiments, the CIRCO framework consists of:

- *CIRCO Schema*: an XML Schema [23,24] model which precisely defines the format of the XML files exchanged among stakeholders' components;
- *CIRCO Java*<sup>2</sup>: an implementation of CIRCO based on the Java<sup>3</sup> programming language to facilitate its adoption and portability.

<sup>2</sup> The documentation is available at the following address:

<http://ims.dei.unipd.it/software/circo/apidoc/>

The source code and the binary code are available at the following address:

<http://ims.dei.unipd.it/software/circo/jar/>

<sup>3</sup> <http://java.sun.com/>

The choice of using an XML-based exchange format is due to the fact that the main other possibility, i.e. to develop a common *Application Program Interface (API)* IR systems have to comply with, presents some issues:

- the experimentation would not be *loosely-coupled*, since all the IR systems would have to be coded with respect to the same API;
- much more complicated solutions would be required for allowing the *distributed* and *asynchronous* running of the experiments, since you would need some kind of middleware for process orchestration and message delivery;
- multiple versions of the API in different languages should be provided to take into account the different technologies used to develop IR system;
- the integration with legacy code could be problematic and require a lot of effort;
- overall, stakeholders would be distracted from their main objective, which is running an experiment and evaluating a system.

## 4 Track Setup

The Grid@CLEF tracks offers a traditional ad-hoc task – see, for example, [1,13] – which makes use of experimental collections developed according to the Cranfield paradigm [5]. This first year task focuses on monolingual retrieval, i.e. querying topics against documents in the same language of the topics, in five European languages: Dutch, English, French, German, and Italian.

The selected languages allow participants to test both romance and germanic languages, as well as languages with word compounding issues. These languages have been extensively studied in the *MultiLingual Information Access (MLIA)* field and, therefore, it will be possible to compare and assess the outcomes of the first year experiments with respect to the existing literature.

This first year track has a twofold goal:

1. to prepare participants' systems to work according to CIRCO framework;
2. to conduct as many experiments as possible, i.e. to put as many dots as possible on the grid.

### 4.1 Test Collections

Grid@CLEF 2009 used the test collection originally developed for the CLEF 2001 and 2002 campaigns [2,3].

**The Documents.** Table 1 reports the document collections which have been used for each of the languages offered for the track.

**Topics.** Topics are structured statements representing information needs. Each topic typically consists of three parts: a brief “title” statement; a one-sentence “description”; a more complex “narrative” specifying the relevance assessment

**Table 1.** Document collections

Language	Collection	Documents	Size (approx.)
Dutch	NRC Handelsblad 1994/95	84,121	291 Mbyte
	Algemeen Dagblad 1994/95	106,484	235 Mbyte
		<b>190,605</b>	<b>526 Mbyte</b>
English	Los Angeles Times 1994	<b>113,005</b>	<b>420 Mbyte</b>
French	Le Monde 1994	44,013	154 Mbyte
	French SDA 1994	43,178	82 Mbyte
		<b>87,191</b>	<b>236 Mbyte</b>
German	Frankfurter Rundschau 1994	139,715	319 Mbyte
	Der Spiegel 1994/95	13,979	61 Mbyte
	German SDA 1994	71,677	140 Mbyte
		<b>225,371</b>	<b>520 Mbyte</b>
Italian	La Stampa 1994	58,051	189 Mbyte
	Italian SDA 1994	50,527	81 Mbyte
		<b>108,578</b>	<b>270 Mbyte</b>

criteria. Topics are prepared in xml format and uniquely identified by means of a *Digital Object Identifier (DOI)*<sup>4</sup>.

In Grid@CLEF 2009, we used 84 out of 100 topics in the set 10.2452/41-AH–10.2452/140-AH originally developed for CLEF 2001 and 2002 since they have relevant documents in all the collections of Table 1.

Figure 4 provides an example of the used topics for all the five languages.

**Relevance Assessment.** The same relevance assessment developed for CLEF 2001 and 2002 have been used; for further information see [2,3].

## 4.2 Result Calculation

Evaluation campaigns such as TREC and CLEF are based on the belief that the effectiveness of IRSs can be objectively evaluated by an analysis of a representative set of sample search results. For this, effectiveness measures are calculated based on the results submitted by the participants and the relevance assessments. Popular measures usually adopted for exercises of this type are Recall and Precision. Details on how they are calculated for CLEF are given in [4]. We used `trec_eval`<sup>5</sup> 8.0 to compute the performance measures.

The individual results for all official Grid@CLEF experiments in CLEF 2009 are given in the Appendices of the CLEF 2009 Working Notes [7].

<sup>4</sup> <http://www.doi.org/>

<sup>5</sup> [http://trec.nist.gov/trec\\_eval](http://trec.nist.gov/trec_eval)



```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<topic>
  <identifier>10.2452/125-AH</identifier>

  <title lang="nl">Gemeenschappelijke Europese munt.</title>
  <title lang="en">European single currency</title>
  <title lang="fr">La monnaie unique européenne</title>
  <title lang="de">Europäische Einheitswährung</title>
  <title lang="it">La moneta unica europea</title>

  <description lang="nl">
    Wat is het geplande tijdschema voor de invoering van de gemeenschappelijke Europese
    munt?
  </description>
  <description lang="en">
    What is the schedule predicted for the European single currency?
  </description>
  <description lang="fr">
    Quelles sont les prévisions pour la mise en place de la monnaie unique européenne?
  </description>
  <description lang="de">
    Wie sieht der Zeitplan für die Einführung einer europäischen Einheitswährung aus?
  </description>
  <description lang="it">
    Qual è il calendario previsto per la moneta unica europea?
  </description>

  <narrative lang="nl">
    De veronderstellingen van politieke en economische persoonlijkheden wat betreft het
    tijdschema waarbinnen men zal komen tot de invoering van een gemeenschappelijke munt voor de
    Europese Unie zijn van belang.
  </narrative>
  <narrative lang="en">
    Speculations by politicians and business figures about a calendar for achieving a
    common currency in the EU are relevant.
  </narrative>
  <narrative lang="fr">
    Les débats animés par des personnalités du monde politique et économique sur le
    calendrier prévisionnel pour la mise en œuvre de la monnaie unique dans l'Union
    Européenne sont pertinents.
  </narrative>
  <narrative lang="de">
    Spekulationen von Vertretern aus Politik und Wirtschaft über einen Zeitplan zur
    Einführung einer gemeinsamen europäischen Währung sind relevant.
  </narrative>
  <narrative lang="it">
    Sono rilevanti le previsioni, da parte di personaggi politici e dell'economia, sul
    calendario delle scadenze per arrivare a una moneta unica europea.
  </narrative>
</topic>

```

Fig. 4. Example of topic <http://direct.dei.unipd.it/10.2452/125-AH>

## 5 Track Outcomes

### 5.1 Participants and Experiments

As shown in Table 2, a total of 2 groups from 2 different countries submitted official results for one or more of the Grid@CLEF 2009 tasks.

Participants were required to submit at least one title+description (“TD”) run per task in order to increase comparability between experiments: all the 18 submitted runs used this combination of topic fields. A breakdown into the separate tasks is shown in Table 3.

The participation in this first year was especially challenging because of the need of modifying existing systems to implement the CIRCO framework. Moreover, it has been challenging also from the computational point of view since, for each component in a IR pipeline, CIRCO could produce XML files that are 50-60 times the size of the original collection; this greatly increased the indexing time and the time needed to submit runs and deliver the corresponding XML files.

**Table 2.** Grid@CLEF 2009 participants

Participant	Institution	Country
chemnitz	Chemnitz University of Technology	Germany
cheshire	U.C.Berkeley	United States

**Table 3.** Breakdown of experiments into tasks and topic languages

Task	# Participants	# Runs
Monolingual Dutch	0	0
Monolingual English	2	6
Monolingual French	2	6
Monolingual German	2	6
Monolingual Italian	0	0
<b>Total</b>		<b>18</b>

## 5.2 Results

Table 4 shows the top runs for each target collection, ordered by mean average precision. The table reports: the short name of the participating group; the mean average precision achieved by the experiment; the DOI of the experiment; and the performance difference between the first and the last participant.

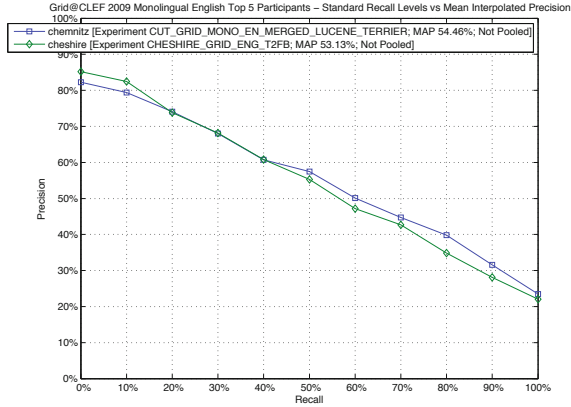
Figure 5 compares the performances of the top participants of the Grid@CLEF monolingual tasks.

## 6 Approaches and Discussion

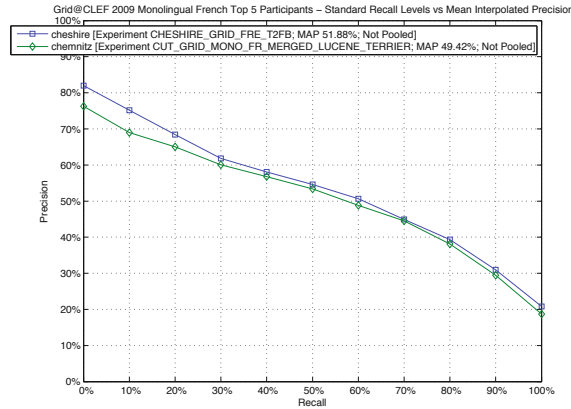
Chemnitz [8] approached the participation in Grid@CLEF into the wider context of the creation of an archive of audiovisual media which can be jointly

**Table 4.** Best entries for the Grid@CLEF tasks

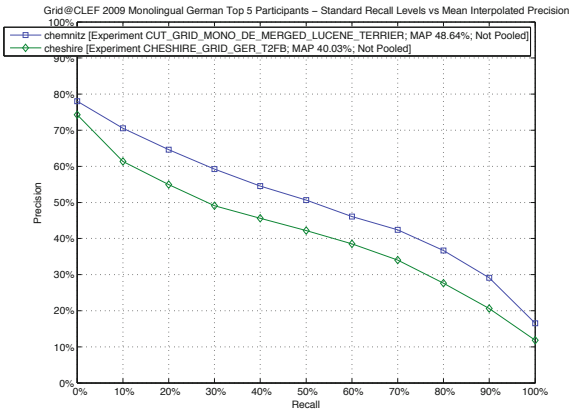
Track	Rank	Participant	Experiment DOI	MAP
English	1st	chemnitz	10.2415/GRIDCLEF-MONO-EN-CLEF2009.CHEMNITZ.CUT_GRID_MONO_EN_MERGED_LUCENE_TERRIER	54.45%
	2nd	cheshire	10.2415/GRIDCLEF-MONO-EN-CLEF2009.CHESHIRE.CHESHIRE_GRID_ENG_T2FB	53.13%
	Difference			2.48%
French	1st	cheshire	10.2415/GRIDCLEF-MONO-FR-CLEF2009.CHESHIRE.CHESHIRE_GRID_FRE_T2FB	51.88%
	2nd	chemnitz	10.2415/GRIDCLEF-MONO-FR-CLEF2009.CHEMNITZ.CUT_GRID_MONO_FR_MERGED_LUCENE_TERRIER	49.42%
	Difference			4.97%
German	1st	chemnitz	10.2415/GRIDCLEF-MONO-DE-CLEF2009.CHEMNITZ.CUT_GRID_MONO_DE_MERGED_LUCENE_TERRIER	48.64%
	2nd	cheshire	10.2415/GRIDCLEF-MONO-DE-CLEF2009.CHESHIRE.CHESHIRE_GRID_GER_T2FB	40.02%
	Difference			21.53%



(a) Monolingual English



(b) Monolingual French



(c) Monolingual German

Fig. 5. Recall-precision graph for Grid@CLEF tasks

used by German TV stations, stores both raw material as well as produced and broadcasted material and needs to be described as comprehensively as possible in order to be easily searchable. In this context, they have developed the Xtrivial system, which aims to be flexible and easily configurable in order to be adjusted to different corpora, multimedia search tasks, and annotation kinds. Chemnitz tested both the vector space model [20,19], as implemented by Lucene<sup>6</sup> and BM25 [17,18], as implemented by Terrier<sup>7</sup>, in combination with Snowball<sup>8</sup> and Savoy's [21] stemmers. They found out that the impact of retrieval techniques are highly depending on the corpus and quite unpredictable and that, even if over they years they have learned how to guess reasonable configurations for their system in order to get good results, there is still the need of "strong rules which let us predict the retrieval quality ... [and] enable us to automatically configure a retrieval engine in accordance to the corpus". This was for them motivation to participate in Grid@CLEF 2009, which represented a first attempt that will allow them to go also in this direction.

Cheshire [14] participated in Grid@CLEF with their Cheshire II system based on logistic regression [6] and their interest was in understanding what happens when you try to separate the processing elements of IR systems and look at their intermediate output, taking this as an opportunity to re-analyse and improve their system, and, possibly, finding a way to incorporate into Cheshire II components of other IR systems for subtasks in which they currently cannot do or cannot do effectively, such as decompounding German words. They also found that "the same algorithms and processing systems can have radically different performance on different collections and query sets". Finally, the participation in Grid@CLEF actually allowed Cheshire to improve their system and to point out some suggestions for the next Grid@CLEF, concerning the support for the creation of multiple indexes according to the structure of a document and specific indexing tasks related to the geographic information retrieval, such as geographic names extraction and geo-referencing.

## Acknowledgements

The authors would like to warmly thank the members of the Grid@CLEF Advisory Committee – Martin Braschler, Chris Buckley, Fredric Gey, Kalervo Järvelin, Noriko Kando, Craig Macdonald, Prasenjit Majumder, Paul McNamee, Teruko Mitamura, Mandar Mitra, Stephen Robertson, and Jacques Savoy – for the useful discussions and suggestions.

The work reported has been partially supported by the TrebleCLEF Coordination Action, within FP7 of the European Commission, Theme ICT-1-4-1 Digital Libraries and Technology Enhanced Learning (Contract 215231).

---

<sup>6</sup> <http://lucene.apache.org/>

<sup>7</sup> <http://ir.dcs.gla.ac.uk/terrier/index.html>

<sup>8</sup> <http://snowball.tartarus.org/>

## References

1. Agirre, E., Di Nunzio, G.M., Ferro, N., Mandl, T., Peters, C.: CLEF 2008: Ad Hoc Track Overview. In: Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandl, T., Peñas, A. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 15–37. Springer, Heidelberg (2009)
2. Braschler, M.: CLEF 2001 – Overview of Results. In: Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.) CLEF 2001. LNCS, vol. 2406, pp. 9–26. Springer, Heidelberg (2002)
3. Braschler, M.: CLEF 2002 – Overview of Results. In: Peters, C., Braschler, M., Gonzalo, J. (eds.) CLEF 2002. LNCS, vol. 2785, pp. 9–27. Springer, Heidelberg (2003)
4. Braschler, M., Peters, C.: CLEF 2003 Methodology and Metrics. In: Peters, C., Gonzalo, J., Braschler, M., Kluck, M. (eds.) CLEF 2003. LNCS, vol. 3237, pp. 7–20. Springer, Heidelberg (2004)
5. Cleverdon, C.W.: The Cranfield Tests on Index Languages Devices. In: Spärck Jones, K., Willett, P. (eds.) Readings in Information Retrieval, pp. 47–60. Morgan Kaufmann Publisher, Inc., San Francisco (1997)
6. Cooper, W.S., Gey, F.C., Dabney, D.P.: Probabilistic Retrieval Based on Staged Logistic Regression. In: Belkin, N.J., Ingwersen, P., Mark Pejtersen, A., Fox, E.A. (eds.) Proc. 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1992), pp. 198–210. ACM Press, New York (1992)
7. Di Nunzio, G.M., Ferro, N.: Appendix D: Results of the Grid@CLEF Track. In: Borri, F., Nardi, A., Peters, C. (eds.) Working Notes for the CLEF 2009 Workshop (2009) (published online)
8. Eibl, M., Kürsten, J.: Putting It All Together: The Xtrieval Framework at Grid@CLEF. In: Peters et al. [15] (2009)
9. Ferro, N.: Specification of the CIRCO Framework, Version 0.10. Technical Report IMS. 2009. CIRCO.0.10, Department of Information Engineering, University of Padua, Italy (2009)
10. Ferro, N., Harman, D.: Dealing with MultiLingual Information Access: Grid Experiments at TrebleCLEF. In: Agosti, M., Esposito, F., Thanos, C. (eds.) Post-proceedings of the Fourth Italian Research Conference on Digital Library Systems (IRCDL 2008), pp. 29–32. ISTI-CNR at Gruppo ALI, Pisa (2008)
11. Ferro, N., Peters, C.: From CLEF to TrebleCLEF: the Evolution of the Cross-Language Evaluation Forum. In: Kando, N., Sugimoto, M. (eds.) Proc. 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access, pp. 577–593. National Institute of Informatics, Tokyo (2008)
12. Ferro, N., Peters, C.: CLEF Ad-hoc: A Perspective on the Evolution of the Cross-Language Evaluation Forum. In: Agosti, F., Esposito, C., Thanos, editors, Post-proceedings of the Fifth Italian Research Conference on Digital Library Systems (IRCDL 2009), pp. 72–79. DELOS Association and Department of Information Engineering of the University of Padua (2009)
13. Ferro, N., Peters, C.: CLEF, Ad Hoc Track Overview: TEL & Persian Tasks. In: Peters et al. [15] (2009)
14. Larson, R.R.: Decomposing Text Processing for Retrieval: Cheshire tries GRID@CLEF. In: Peters et al. [15]

15. Peters, C., Di Nunzio, G.M., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A., Roda, G. (eds.): Multilingual Information Access Evaluation Vol. I Text Retrieval Experiments – Tenth Workshop of the Cross-Language Evaluation Forum (CLEF 2009). Revised Selected Papers. LNCS. Springer, Heidelberg (2010)
16. Robertson, S.E.: The methodology of information retrieval experiment. In: Spärck Jones, K. (ed.) *Information Retrieval Experiment*, pp. 9–31. Butterworths, London (1981)
17. Robertson, S.E., Spärck Jones, K.: Relevance Weighting of Search Terms. *Journal of the American Society for Information Science (JASIS)* 27(3), 129–146 (1976)
18. Robertson, S.E., Walker, S., Beaulieu, M.: Experimentation as a way of life: Okapi at TREC. *Information Processing & Management* 36(1), 95–108 (2000)
19. Salton, G., Buckley, C.: Term-weighting Approaches in Automatic Text Retrieval. *Information Processing & Management* 24(5), 513–523 (1988)
20. Salton, G., Wong, A., Yang, C.S.: A Vector Space Model for Automatic Indexing. *Communications of the ACM (CACM)* 18(11), 613–620 (1975)
21. Savoy, J.: A Stemming Procedure and Stopword List for General French Corpora. *Journal of the American Society for Information Science (JASIS)* 50(10), 944–952 (1999)
22. Savoy, J.: Report on CLEF-2001 Experiments: Effective Combined Query-Translation Approach. In: Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.) *CLEF 2001*. LNCS, vol. 2406, pp. 27–43. Springer, Heidelberg (2002)
23. W3C. XML Schema Part 1: Structures – W3C Recommendation (October 28, 2004), <http://www.w3.org/TR/xmlschema-1/>
24. W3C. XML Schema Part 2: Datatypes – W3C Recommendation (October 28, 2004), <http://www.w3.org/TR/xmlschema-2/>
25. W3C. Extensible Markup Language (XML) 1.0 (Fifth Edition) – W3C Recommendation (November 26, 2008), <http://www.w3.org/TR/xml/>