

# Do Expressive Geographic Queries Lead to Improvement in Retrieval Effectiveness?

Damien Palacio<sup>1,3</sup>, Christian Sallaberry<sup>1</sup>, Guillaume Cabanac<sup>2</sup>, Gilles Hubert<sup>2</sup>, and Mauro Gaio<sup>1</sup>

<sup>1</sup> Université de Pau et des Pays de l'Adour, LIUPPA ÉA 3000, France

<sup>2</sup> Université de Toulouse, IRIT UMR 5505 CNRS, France

<sup>3</sup> GIS, Geographisches Institut, Universität Zürich, Switzerland

damien.palacio@geo.uzh.ch  
{christian.sallaberry, mauro.gαιο}@univ-pau.fr  
{guillaume.cabanac, gilles.hubert}@irit.fr

**Abstract.** In an information retrieval (IR) context, users usually issue queries with few terms and no operators (e.g., and, or, +). However, most of users' information needs involve more expressiveness (e.g., '*Potato famine in Ireland, but not in Cork*'). Our work deals with this category of queries that may be processed by geographic IR (GIR) systems to parse digital libraries according to spatial, temporal, and topical criteria. We propose a GIR framework that supports expressive queries and aggregates results of a multi-criteria search. We also conduct experiments to verify that this approach improves the effectiveness of such search engines (improvement of 27% for topical criteria only, and of 54% for spatial and temporal criteria).

## 1 Introduction

Several studies (Silverstein et al. 1999; Spink et al. 2001; White and Morris 2007) on popular search engines, such as Altavista, Excite, Google, MSN Search, or Yahoo! showed that the users of web-based search engines typically issue queries with few terms (between 2 and 3) and rarely use operators (up to 20%). In the same time, Lucas and Topi (Lucas and Topi 2002) analyzed the effects of term and operator usage on the relevance of search results. They showed that queries from experts involve more terms and operators (up to 4 times more), without yielding improvement in effectiveness in most cases.

---

Preprint of :

Damien Palacio, Christian Sallaberry, Guillaume Cabanac, Gilles Hubert, Mauro Gaio (2012). *Do Expressive Geographic Queries Lead to Improvement in Retrieval Effectiveness?* Proceedings of the International Conference on Geographic Information Sciences (AGILE 2012), Jérôme Gensel, Didier Josselin, Danny Vandenbroucke (Eds.), Springer, LNGC, p. 267-286.

[http://dx.doi.org/10.1007/978-3-642-29063-3\\_15](http://dx.doi.org/10.1007/978-3-642-29063-3_15)

They pointed out two groups of factors to explain most of the variations in the relevance of search results: users may fail using the appropriate term or they fail using the relevant operator when issuing their query.

In the context of digital humanities, our work is focused on digital libraries (DLs). When compared to web information retrieval (IR), IR in DLs has two specificities. (i) Corpora contents refer to domain specific topics (e.g., wild-life, flora, art, sports). As a consequence, users' needs and vocabulary are usually more specific than in common corpora. (ii) Users are often accustomed to query operators (e.g., librarians, scholars) (Aula et al. 2010). Let us consider the example of an expressive search '*Potato famine in Ireland after mid-19th century; prioritizing documents evoking the Connacht province; excluding those evoking Cork or any other location included in Cork city.*' The more expressive the query, the more criteria and corresponding operators are used. We consider the usual acceptance that Geographic Information gathers three dimensions, namely spatial, temporal, and topical (Usery 1996). Fig. 1 illustrates these dimensions, as well as the variety of criteria within each dimension often used when searching for information in DLs. This is reinforced by Larson (Larson 2009) when he demonstrates the effectiveness of Geographic Information Retrieval (GIR) methods in DLs.

In (Palacio et al. 2010) we introduced a system supporting the basic part of such queries (regular part of the query in Fig. 1) with matching operators dedicated to each dimension. However, during the query formulation process, the user often introduces more expressiveness with modal operators such as *May*, *Should*, *Must*, *Must not*, to give priority to a criterion or to use a criterion as an exclusion key (expressive part of the query in Fig. 1). Therefore, in this paper, we propose the CMRP (Criterion, Matching, Requirement, Preference) multi-criteria IR framework to explore DLs. It supports the association of roles (*Exclusion*, *Demotion*, *Regular*, *Promotion*, *Necessity*, or *Target*) to any criterion in order to extend the expressiveness of queries (regular and expressive parts of the query in Fig. 1). We hypothesize that the association of roles to query criteria will improve result effectiveness. The CMRP framework is implemented within the geographic IR engine called PIV<sup>3</sup>, for 'Virtual Itineraries in the Pyrenees Mountains<sup>1</sup>'.

We developed spatial and temporal IR systems (IRSs) dedicated to unstructured textual documents (Gaio et al. 2008; Le Parc-Lacayrelle et al. 2007). Now, the challenge consists in aggregating spatial, temporal, and topical IRSs within a single IR framework keeping in mind that such a framework may further integrate IRSs dedicated to documents of different formats. Moreover, the aggregation approach has to support expressive roles associated with any criterion.

---

<sup>1</sup> The PIV project is funded by the Pau City Council and the MIDR multimedia library.

The paper is organized as follows. In Sect. 2, we review the literature related to multi-criteria search and GIR systems, as they may benefit from our framework. In Sect. 3, we describe the CMRP framework and the PIV<sup>3</sup> search engine: it presents the core components of PIV<sup>3</sup>—spatial, temporal, and topical IRSs—and their federation within PIV<sup>3</sup>. In Sect. 4, we report evaluations through experiments complying with the evaluation framework proposed in (Palacio et al. 2010). Finally, Sect. 5 concludes the paper and outlines research directions.

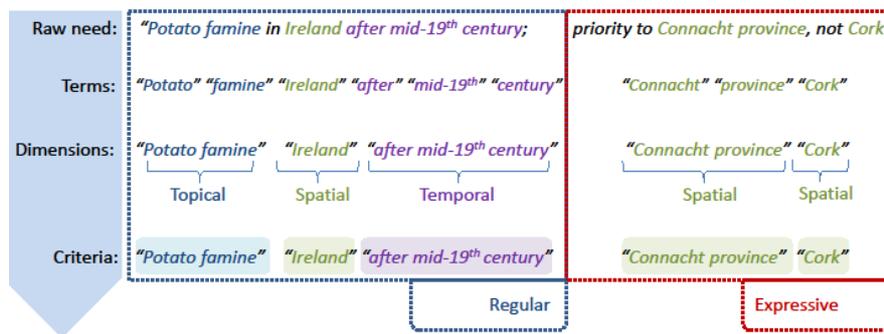


Fig. 1. Raw query as issued by the user, and segmentation into dimensions and criteria

## 2 Related Work: Multi-Criteria Search and Geographic IR

For classical IR, more expressive queries lead to better results. Indeed, Kelly and Fu (Kelly and Fu 2007) showed a strong relationship between query expansion (with additional information from users) and performance.

Croft and Harabagiu (Croft and Harabagiu 2000) promoted the combination of different text representations and search strategies as a technique for improving the effectiveness of information retrieval. They described three categories of approaches: the combination of representations given as input for retrieval algorithms, the combination of retrieval algorithms, and the combination of search system outputs. Fox and Shaw's (Fox and Shaw 1993), as well as Fernandez et al.'s (Fernández et al. 2006) score-based aggregation models may be classified in the third category. Both propose to normalize the source scores to a common distribution before being combined.

Therefore, we make the assumption that various search criteria plus advanced roles specifying whether a criterion is required to retrieve a document or even a criterion is used to promote, demote, or reject a document will improve descriptions of information needs and result effectiveness. In this section, we first focus on recent work dedicated to the aggregation of result lists issued from multi-criteria IR. Then, we briefly describe work related to multi-

criteria IR dedicated to the handling of spatial and temporal dimensions of documents.

## 2.1 Multi-Criteria Information Retrieval

In the context of multi-criteria IR, Farah and Vanderpooten (Farah and Vanderpooten 2008) define result aggregation as a process that ranks documents by combining document retrieval status values (RSVs), also known as ‘score,’ associated with each criterion. Fig. 2 illustrates the principle of multi-criteria retrieval. A multi-criteria query conveys the user’s need. The search engine parses the corpus in order to match the query and document contents. A result list  $L_i$  comprising presumably relevant documents with corresponding RSVs is produced for each criterion. Then, the search engine aggregates those result lists to produce the final result  $L$  which is displayed to the user.

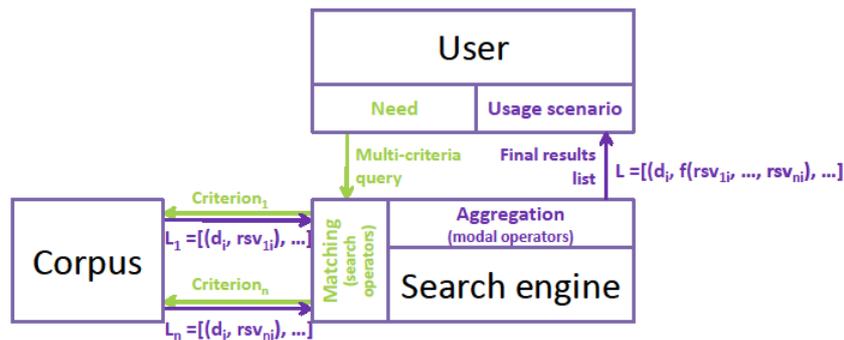


Fig. 2. Principle of multi-criteria retrieval

Farah and Vanderpooten (Farah and Vanderpooten 2008) consider three classes of aggregation functions and describe them as follows:

**Totally Compensatory Logic.** This consists in computing a single RSV using functions such as the weighted sum. For such functions, a low score for one criterion can be compensated by one or several high scores for other criteria. So, this kind of approach does not handle any required criterion. Prominent compensatory functions are Borda Count (de Borda 1781), CombMAX, CombMNZ, and CombSUM (Fox and Shaw 1993)

**Non-Compensatory Logic.** In this case, aggregation is mainly based on the value of one criterion, such as the lowest score of all criteria. The remaining criteria are only used to discriminate between documents with similar scores. A clear weakness of this class of operators is that a large part of the scores is ignored or plays a minor role. CombMIN (Fox and Shaw 1993) and Prioritized And (Costa Pereira et al. 2009) are functions of the non-compensatory class.

**Partially Compensatory Logic.** Farah and Vanderpooten (Farah and Vanderpooten 2008) propose to combine any set of criteria while taking into consideration the imprecision underlying the query formulation process. In a similar way, da Costa Pereira et al. (Costa Pereira et al. 2009) propose the Prioritized Scoring Model (PSM) aggregation function giving the most important weight to the first criterion and decreasing weights to remaining criteria. Considering the important need of expressiveness for queries in the context of DLs, we make the assumption that partially compensatory functions supporting roles (e.g., exclusion, demotion, promotion, necessity) associated to query criteria may improve result effectiveness.

GIR is a domain where complex queries are issued, especially when searching for information in DLs (Larson 2009). Moreover, the ratio of geographic queries submitted to usual search engines is about 15% according to studies described in (Gan et al. 2008; Jones et al. 2008; Sanderson and Kohler 2004), and may increase in the context of DLs. Consequently, we propose a brief review of GIR frameworks in the next section.

## 2.2 Frameworks for Geographic Multi-Criteria IR

Among specialized fields of IR, Geographic IR (GIR) is concerned with multi-criteria queries. Indeed, a piece of geographic information is generally described as a set of three dimensions: spatial, temporal, and topical (Userly 1996). As a result, lots of GIR systems support multi-criteria geographic queries to handle those three dimensions.

Concerning non-compensatory aggregation, CITER (Pfoser et al. 2009) relies on a ‘parallel filtering’ approach: this system targets simultaneously and separately each dimension, and then combines the different result lists by computing their intersection. Because this parallel filtering operation is based on the Set theory, documents relevance is binary. As a result, the retrieved documents are not ranked.

Concerning compensatory aggregation, DIGMAP (Martins et al. 2007) relies on the ‘linear interpolation’ approach (e.g., arithmetic mean). Nevertheless, such aggregation models may be biased as they combine results returned by different IRSs characterized by specific information representation and score computation approaches.

To the best of our knowledge, all GIR systems use a non-compensatory or a compensatory aggregation; none use partially compensatory aggregation. For a more detailed presentation of current GIR systems we refer the reader to (Palacio et al. 2010).

Systems tackling several dimensions for search purpose, like GIR systems, usually federate mono-dimensional IR systems, each one supporting its own domain-specific matching operators. Such multidimensional systems may benefit from partially compensatory aggregation functions supporting roles

for refining the relative importance of criteria (Fig. 3). The next section presents the proposed CMRP framework, which provides a solution to design and process expressive queries.

### 3 CMRP: a Partially Compensatory Aggregation-Based IR Framework

CMRP is a GIR framework dedicated to textual corpora. It involves process flows dedicated to spatial and temporal information recognition and interpretation (Gaio et al. 2008; Le Parc-Lacayrelle et al. 2007). It builds several levels of index describing detailed spatial and temporal extracted entities, as well as larger points of view supported by spatial and temporal tile-based generalization approaches (Palacio et al. 2011). Though, it enables expressive queries, for which constraints can be attached to any criterion.

Let us consider again the example of the expressive search ‘*Potato famine in Ireland after mid-19th century; prioritizing documents evoking the Connacht province; excluding those evoking Cork or any other location included in Cork city.*’ Fig. 3 illustrates the variety of matching operators and roles. Matching operators are dimension-dependent (e.g., spatial, temporal, topical) whereas roles correspond to the expressiveness supported by the CMRP framework. Fig. 3 details dimensions and criteria of the query introduced in Fig. 1: predefined roles and/or personalized modal operators are associated with each criterion.

As users have sophisticated needs, they may formulate expressive queries in two different ways. Regular users assign one of the various proposed roles (detailed in Sect. 3.1) to a criterion (e.g., *Must* is a mandatory role, *Must not* is an exclusion role, *May* is an enhancement one).



Fig. 3. Criterion roles and modal operators for an expressive query composition

Advanced users may directly combine the +, -, ^ modal operators (Fig. 3) to refine these roles (e.g., the ^ operator gradually penalizes or boosts a crite-

tion—weighted between  $[-1, 1]$ ). The next section introduces the CMRP framework.

### 3.1 Proposed CMRP Framework

We designed a framework to handle expressive multi-criteria queries. Let  $D$  denote the set of documents. This is based on the quadruplet  $(C, M, R, P)$  defined as follows:

- $C = (c_1, \dots, c_n)$ , where  $c_i$  is a criterion expressed in the query.
- $M = (m_1, \dots, m_n)$ , where  $m_i: C \times D \rightarrow [0, 1]$ .  $m_i$  refers to an external matching function used to compare any document  $d \in D$  with criteria  $c_i$ . Various operators may be available in the CMRP framework, such as intersection ( $\cap$ ), equality ( $=$ ), inclusion ( $\subseteq$ ), proximity ( $\sim$ ). They correspond to the functions supported by the invoked IRSs.
- $R = (r_1, \dots, r_n)$ , where requirement  $r_i$  may be mandatory (+), neutral, (N) or excluded (-).
- $P = (p_1, \dots, p_n)$ , with  $p_i \in \mathbb{R}$ , where preference  $p_i$  allows a user to weight criterion  $c_i$  according to the corresponding level of importance he/she wants.

Table 1 illustrates the expressive query given in Fig. 3. In this example, notice that the temporal criterion is not mandatory, but boosted whenever it is encountered in a document. In addition, the topical criterion is mandatory and boosted. Notice that the preferences are chosen by the user.

**Table 1.** Expressive query and result combination

Quadruplet	Criterion				
	Topical	Spatial	Temporal	Spatial	Spatial
C	potato famine	Ireland	1850-1900	Connacht	Cork
M	=	$\cap$	$\cap$	=	$\subseteq$
R	+	+	N	N	-
P	0.5	0	0.7	0.7	0

We identified six predefined roles (Fig. 4) that may be associated with a criterion: exclusion (negative filtering), demotion, regular, promotion, necessity (positive filtering), and target. Table 2 and Fig. 4 illustrate these different scenarios that we detail in this section.

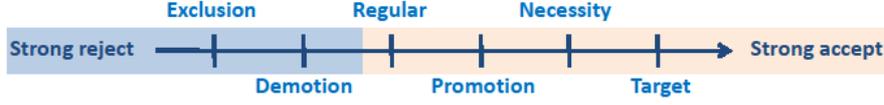


Fig. 4. Six predefined possible roles refining the expressivity of a search criterion

Table 2. Mapping of criteria roles to requirements and preferences

Criteria roles	Requirements			Preferences		
	-	N	+	$\mathbb{R}_-^*$	N	$\mathbb{R}_+^*$
Exclusion	✓				✓	
Demotion		✓		✓		
Regular		✓			✓	
Promotion		✓				✓
Necessity			✓		✓	
Target			✓			✓

**A - Exclusion.** A user can decide to exclude one criterion. This criterion is used to reject documents and so it is not necessary to give it a preference.

**B - Demotion.** A user can choose to depreciate one criterion. Documents that contain this criterion are not dropped, but rather their global score are decreased.

**C - Regular.** A user can use a criterion without setting any preference or requirement.

**D - Promotion.** A user can decide to include an optional criterion. If this criterion is present, the document is more interesting, otherwise it is without consequences. Thus, a weight is associated with this criterion to influence the calculation of the global score.

**F - Necessity.** A user can decide to use a mandatory criterion. So, this criterion will not influence the global score calculation. This criterion must be present, but no preference is set.

**E - Target.** One criterion (or several) can be defined as the search goal. So, it must be present and it influences the global score calculation. If all criteria are defined as targets with similar weights, they are all equally relevant for the search.

Note that some combinations are incompatible (for example, an exclusion (-) requirement and a positive ( $\mathbb{R}_+^*$ ) preference). As a result, it is mandatory to check the validity of each combination. Such combination scenarios are not allowed by the framework and are rejected.

$$RSV(d_i) = \begin{cases} 0 & \text{if } d_i \sim R \\ \frac{1}{\sum_{i=1}^n p_i | p_i > 0} \cdot \sum_{i=1}^n p_i \cdot m_i(d_i, c_i) & \text{else} \end{cases} \quad (1)$$

The CMRP approach allows one to improve the expressiveness of each criterion in the query. The final score of a document is computed by a partially compensatory aggregation function (Eq. 1). This combines scores so that each criterion has an effect on the final score. Notice that, if a mandatory criterion is missing or if an excluded one is present in a document (denoted as  $d_i \not\sim R$ ), this document is not retrieved (i.e.,  $RSV(d_i) = 0$ ). As a result, a mandatory or excluded criterion can not be compensated.

To sum up, we introduced a new search process that extends those applied in criteria aggregation (Farah and Vanderpooten 2008) in order to process multi-criteria queries. CMRP has several advantages: (i) For each criterion, matching functions depend on dimension-dedicated IRSs. (ii) For each criterion, expressiveness (i.e., requirement or preference) is independent of the invoked IRS. (iii) For a query, result aggregation is partially compensatory: the equation for score computation takes into account the user's requirement and preference weights associated with each criterion. Now we present this framework implementation within the PIV<sup>3</sup> meta-engine that handles multi-criteria queries.

### 3.2 PIV<sup>3</sup> Meta-Engine Implementing CMRP Framework

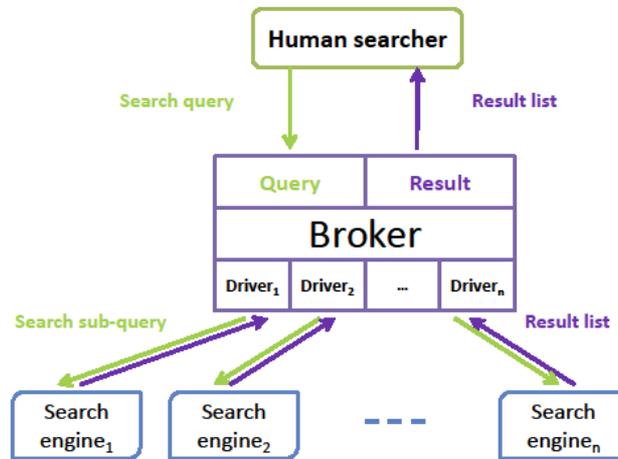
The CMRP framework can be implemented on top of a single search engine, or with several search engines federated by a meta-search engine. Building on Rasolofo et al. (Rasolofo et al. 2003), we followed the second approach to implement the CMRP framework within the PIV<sup>3</sup> meta-search engine.

The PIV project aims to manage a digitized collection of documents published in the 19th century about the French Pyrenees Mountains. Following recent work on Digital Libraries (Liesaputra et al. 2009), the main goal of the PIV project is to help users find accurate information inside books. It intends to overcome usual IRS limitations regarding geographic information management. Thus, we designed three process chains for spatial, temporal, and topical information indexing (Gaio et al. 2008; Le Parc-Lacayrelle et al. 2007). The generated indexes allow the retrieval of document units along with relevance scores consistent with any spatial, temporal, and topical dimension.

Fig. 5 describes a meta-search engine as a broker, which splits a query into sub-queries dedicated to the targeted search engine. It also supports a result list aggregation process that produces a single list of results by merging several result lists.

The PIV<sup>3</sup> meta-search engine includes an instance of the CMRP framework (Sect. 3.1) within the global architecture described in Fig. 5. The PIV<sup>3</sup> drivers feature three primary search engines: PIV\_spatial, PIV\_temporal, and PIV\_topical (based on the Terrier (Ounis et al. 2005) search engine). Spatial,

temporal, and term-based specific matching functions are declared in these drivers. As a result, PIV<sup>3</sup> federates three mono-dimensional IRSs. As it embeds dimension-specific operators, it splits a multi-criteria query into sub-queries and forwards them to the corresponding IRSs. Moreover, PIV<sup>3</sup> supports the CMRP model to extend the expressiveness of the different criteria as defined in Sect. 3.1. It computes a partially compensatory aggregation to build a merged list of results comprising ranked documents.



**Fig. 5.** A meta-search engine architecture according to Rasolofo et al. (Rasolofo et al. 2003)

Another important feature of this approach lies in the generalized information representations and retrieval models supported by PIV\_spatial and PIV\_temporal primary IRSs (Palacio et al. 2010). Each IRS supports a tile-based generalization of spatial or temporal information. Such an approach features a generalization resulting in an homogeneous and comparable representation of the various dimensions of geographic information. It is related to well-known IR models based on term lemmatization/truncation and weighting schemes (e.g., term frequency in the Vector Space Model). The final scores are not biased since representations and scores were generalized prior to the combination process. We refer the reader to (Palacio et al. 2010) for further explanations on this subject.

In the next section, we evaluate the CMRP framework implemented in the PIV<sup>3</sup> meta-search engine and compare it to CombMNZ and PSM aggregation functions.

## 4 Experiments

We aim at evaluating a search engine that handles expressive multi-criteria queries in a geographical context. The problem is, as explained in (Palacio et al. 2010), that there is no test collection covering all geographic dimensions (spatial, temporal, and topical). So, we propose to evaluate the CMRP framework, implemented in PIV<sup>3</sup>, according to two case studies:

- **TREC test collection** (Voorhees and Harman 1999). Multi-criteria queries are submitted to Terrier search engine (with TF·IDF IR model) whereas operators are managed by PIV<sup>3</sup>. Thus, the CMRP framework can be used to extend topical search engines that do not support advanced operators or that support only some of them.
- **MIDR\_2010 test collection** (Palacio et al. 2010). Multi-criteria queries are submitted to PIV\_Spatial, PIV\_Temporal and Terrier search engines that are federated within PIV<sup>3</sup>.

During these experiments, queries were built as follows: (i) Criteria were listed explicitly by the user; (ii) then, expressiveness was specified upon each criterion (matching operators/modal roles) by an automated topic variant generator (case study 1) or by the user (case study 2).

### 4.1 Topical multi-criteria IR

As we did in (Hubert et al. 2011), we conducted experiments on the TREC-8 (Voorhees and Harman 1999) standard test collection. The TREC-8 corpus of newspaper articles corresponds to the kind of document units (e.g., newspaper or travelogue paragraphs) that the PIV<sup>3</sup> engine is expected to retrieve. Moreover, it provides  $n = 50$  topics covering various subjects, allowing us to realize significance testing. Each topic is worded as: (i) A *title*: a sequence of two or three terms (in general) that would be submitted as a query to a search engine by an average user. (ii) A *description*: a few sentences describing the user's information needs in plain text. (iii) A *narrative*: a longer text than the description, which gives insights into the intent of the user, and unambiguously states what information is relevant or irrelevant for the searcher.

We demonstrated the benefits of query operators (+ and ^) on regular topics in (Hubert et al. 2011). Here, because topics contain few terms and none that can be associated with exclusion or demotion roles, we propose to use the narrative part to extend each topic title with new terms. These terms are split into two categories: (i) positive terms are associated randomly with necessity (+), promotion ( $\wedge\mathbb{R}_+^*$ ) and target roles; (ii) negative terms are associated randomly with exclusion (-) and demotion ( $\wedge\mathbb{R}_-^*$ ) roles.

Therefore, the experiment evaluates the three following scenarios:

1. *Regular topic*. We consider this as the baseline using the original titles of the TREC-8 test collection.
2. *Extended topic*. This is comprised of the positive terms only from title and some manually extracted from the narrative. No operator is associated with them.
3. *Expressive topic*. This is composed of the terms composing both positive and negative categories. Operators are randomly associated with these terms.

Within the *Expressive topic* scenario, we intend to check that any topic can be rewritten with operators, such that it leads to more accurate search results.

Then, as explained in (Hubert et al. 2011), a *Topic Variant Generator* is used to generate up to 6,561 variants (i.e., 4 distinct terms) per topic, due to all the possible combinations of the terms with potential roles. For the 50 topics we have 188,973 variants. The following example (Table 3) illustrates some variants generated for the *Hurricanes but not damages* topic.

Generating variants for  $t_p$  positive terms leads to  $9^{t_p}$  possibilities ( $+t_{p1}$ ,  $t_{p1}^{+0.25}$ ,  $t_{p1}^{+0.5}$ ,  $t_{p1}^{+0.75}$ ,  $t_{p1}^{+1}$ ,  $+t_{p1}^{-0.25}$ ,  $+t_{p1}^{-0.5}$ ,  $+t_{p1}^{-0.75}$ ,  $+t_{p1}^{-1}$ ) and for  $t_n$  negative terms leads to  $6^{t_n}$  possibilities ( $\emptyset$ ,  $-t_{n1}$ ,  $t_{n1}^{-0.25}$ ,  $t_{n1}^{-0.5}$ ,  $t_{n1}^{-0.75}$ ,  $t_{n1}^{-1}$ ). Therefore, the total of variants is  $9^{t_p} \times 6^{t_n}$ .

**Table 3.** Excerpt of a topic variants generated with random combinations: positive terms associated with necessity, promotion and target roles and negative terms associated with exclusion and demotion roles

Variant #	Query variants	
1	+hurricane	$\emptyset$
2	+hurricane	-damage
3	+hurricane	damage $^{-0.25}$
4	+hurricane	damage $^{-0.5}$
.	.	.
.	.	.
54	+hurricane $^{+1}$	damage $^{-1}$

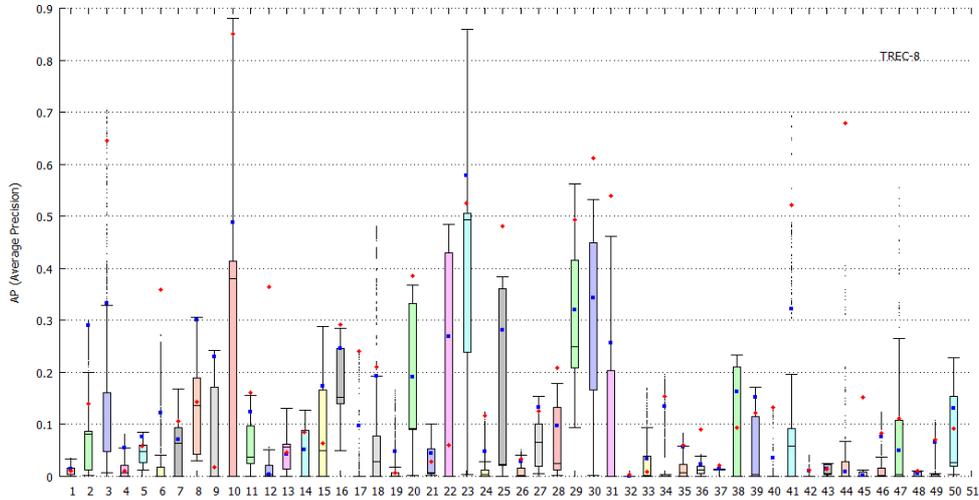
We queried PIV<sup>3</sup> with each of the *regular*, *expressive* and *extended topics* corresponding to the 50 TREC-8 topics. To study the distribution of effectiveness data values we use ‘boxplot’ visualization (Tukey 1977; Williamson et al. 1989) with the Gnuplot software<sup>2</sup>. A boxplot (a.k.a. box-and-whisker diagram) summarizes several descriptive statistics. The interquartile range (IQR) spans the lower quartile to the upper quartile. The middle 50% of the ranked data lies in the IQR. It is represented as a *box* (central rectangle), which shows the spread of data values. The median is shown as a *segment* inside the box.

<sup>2</sup> <http://www.gnuplot.info/>

This is the middle half of the data values, and allows one to assess the symmetry of the distribution. The *whiskers* extend from the ends of the box to the most distant value lying within  $1.5 \times \text{IQR}$ . Larger and lower values are considered as outliers; these are plotted with *black circles*.

Fig. 6 shows effectiveness results for the three groups of topics. These results can be organized in 4 categories:

- Expressive topics improve upon regular topics but extended topics are quite similar (less than 5% of difference between results of expressive and extended topics) for 6% of all cases. Regular topic expansion with new terms is enough (e.g., queries number 2, 8, and 9; Fig. 6).
- Expressive topics overcome extended topics, as well as regular ones for 26% of all cases. Several (one or more) of the randomly produced term/operator combinations improve retrieval results (e.g., queries number 33, 38, and 39; Fig. 6).
- Expressive topics improve upon regular topics, when extended topics do not overcome regular ones, for 32% of all cases. Expansion based on new terms does not improve the baseline. However, when associated with convenient operators, these terms improve retrieval results (e.g., queries number 10, 29, and 34; Fig. 6).
- Expressive topics overcome extended topics but not regular ones for 32% of all cases. Expansion based on new term addition is not convenient although operators improve significantly their retrieval results (e.g., queries number 6, 30, and 31; Fig. 6).



**Fig. 6.** Average Precision (AP) boxplots showing potential for improvement with expressive topics on 50 queries (TREC-8). Diamonds  $\blacklozenge$  show the AP of regular topics, diamonds  $\blacksquare$  show the AP of extended topics and the higher positions of boxplots indicate the AP corresponding to the best topic variant.

As shown in Table 4, Extended topics reduce results precision ( $-30.61\%$ ), whereas expressive topics improve results ( $+26.63\%$ ), and it is statistically significant ( $t$ -test with  $p < 0.05$ ). If we compare expressive topics and extended topics (considering that expressive topics add roles to extended ones), the improvement is even bigger ( $+82.98\%$ ).

Several reasons can explain why extended topics are worse than the baseline: (1) the original query is already ideal so extension only brings noise; (2) the original query can be improved, but adding keywords with similar importance (totally compensatory approach) only brings noise. Nevertheless, these keywords associated with modal operators bring suited expressiveness and improve results. It shows the difficulty to choose the good terms.

**Table 4.** Experiment comparing different scenarios

Topics	MAP	Improvement (%)	$t$ -test ( $p$ -value)
Regular (baseline)	0.1970	0.00	0.0000
Extended	0.1367	-30.61	0.0097
Expressive	0.2502	26.96	0.0098

Overall, notice that, among all the variants of an expressive topic, the best one implies different promotion roles (92% of all cases) and a necessity role

(80% of all cases). Moreover, 50% of topics involve exclusion or demotion criteria and about 40% of them contribute to topic improvement.

Therefore, when terms of an expressive topic are well chosen there always exists a multi-criteria query that may improve the retrieval results. As geographical queries often involve two or three dimensions, we believe they are more suited to such a multi-criteria querying approach: specific spatial and temporal search criteria are easier to identify than term-based criteria and corresponding roles are also easier to specify. The next experiment evaluates PIV<sup>3</sup> engine with a geographical test collection. So we verify that with a better suited context, PIV<sup>3</sup> will improve effectiveness in a larger extent.

## 4.2 Geographic Multi-criteria IR

We used the MIDR\_2010 cultural heritage test collection (Palacio et al. 2010) (available on the PIV project website<sup>3</sup>), whose corpus represents 5,645 paragraphs extracted from 11 books published between the 18th and 20th centuries. We manually designed 10 expressive queries targeting geographic items present in this DL (e.g., Fig. 1 query: *'Potato famine in Ireland after mid-19th century; priority to Connacht province, not Cork'*). Moreover, we measured effectiveness of the tested IRSs with respect to *Mean Average Precision* (MAP) (Manning et al. 2008). Let us note that in (Palacio et al. 2010) we only experimented the CombMNZ totally compensatory aggregation function.

Here, we first compare a full compensatory aggregation function (CombMNZ) (Fox and Shaw 1993) to the partially compensatory aggregation function, 'Prioritized Scoring Model' (PSM), proposed by da Costa Pereira et al. (Costa Pereira et al. 2009) (Sect. 2.1). The main advantage of the PSM function is that a user just needs to specify the order of criteria since decreasing weights of each criterion are calculated automatically. Although PSM is intended to improve personalized search, it is also appropriate for dealing with expressive queries. It is thus interesting to compare its effectiveness to the CombMNZ function and to the CMRP model designed to enhance search with expressive queries. With the PSM approach, the query of Fig. 1 can be expressed as: 'potato famine > Ireland > 1850-1900 > Connacht'. 'not Cork' cannot be expressed because exclusion is not supported in PSM. Nevertheless, in this first experiment, as shown in Table 5, PSM yields weaker results than the CombMNZ baseline. This difference is not statistically significant however (*t*-test with  $p > 0.05$ ). PSM is a partially compensatory aggregation function however it does not support expressive queries involving advanced roles like necessity, target, exclusion, demotion. Therefore, in such a context, a full compensatory aggregation function returns better results.

---

<sup>3</sup> <http://t2i.univ-pau.fr/MIDR/>

**Table 5.** First experiment comparing aggregation approaches

Models	MAP	Improvement (%)	<i>t</i> -test ( <i>p</i> -value)
CombMNZ (baseline) (Fox and Shaw 1993)	0.1658	0.00	0.00
Prioritized Scoring Model (PSM) (Costa Pereira et al. 2009)	0.1034	-37.64	0.18

**Table 6.** Second experiment comparing aggregation approaches

Models	MAP	Improvement (%)	<i>t</i> -test ( <i>p</i> -value)
CombMNZ (baseline) (Fox and Shaw 1993)	0.1658	0.00	0.00
CMRP (PIV <sup>3</sup> )	0.2556	+54.16	0.10

Second, we compare the CMRP framework based on a partially compensatory aggregation and implemented by PIV<sup>3</sup> to the CombMNZ baseline with the same queries. As shown in Table 6, the CMRP framework yields an improved effectiveness by 54.16%, which is not statistically significant however ( $p > 0.05$ ). We need to continue our experiment (more than 10 queries) to be able to rely on statistical testing. The observed difference may be due to the fact that CMRP allows users to refine roles for each criterion, unlike CombMNZ. Indeed, CombMNZ does not support requirement or preference modal operators. As the CMRP framework allows users to refine several roles, it enables enhanced expressiveness for queries.

Both experiment settings rely on documents retrieved with quite low average precisions (MAP). This is due to several restrictive topics in the test collection. However, systems participating in TREC have similar results (Voorhees and Harman 1999).

To sum up, the first experiment of this section showed that it is better to use a full compensatory aggregation function (e.g., CombMNZ) than to use a function that does not take into account the expected expressiveness. The second experiment showed that CMRP framework yields better results (+54.16%) as it provides the user with a way to express advanced roles for each criterion.

## 5 Conclusion and Future Work

In this paper, we considered expressive user needs formulated in various contexts, such as DLs. Users spend efforts and time creating expressive que-

ries. This process leads to multi-criteria queries, where each criterion may be refined by a role describing its importance and semantics within the query.

We proposed the CMRP framework for handling preferences and requirements in search queries. Requirements are achieved by mandatory, exclusion, or neutral modal operators whereas preferences are achieved by positive or negative amounts associated to a boosting modal operator in order to promote or demote a document. CMRP integrates them in a partially compensatory aggregation function computed to retrieve a single result list from different criteria-based result lists.

As a proof of concept, we designed the PIV<sup>3</sup> GIR meta-search engine, which implements the CMRP framework. This system takes spatial, temporal, and topical dimensions into account, so as to provide more expressiveness to users who express their information needs into queries. Each domain-specific IRS supports spatial, temporal, or topical matching functions whereas the meta-search engine allows the end-user to associate any criteria with expressive roles. A first experiment based on TREC-8 topical queries shows that PIV<sup>3</sup> significantly improves search results (+26.96%). A second experiment based on geographical queries, and the cultural heritage test collection MIDR\_2010 shows that PIV<sup>3</sup> GIRS outperforms (+54.16%) CombMNZ (Fox and Shaw 1993).

These corpora and dedicated IR systems deal with unstructured textual documents. However, our framework may integrate other IRSs dedicated to geographic information contained in documents of different formats: image, video, or sound. PIV<sup>3</sup> architecture may federate such various IR systems; the only constraint is to build the corresponding drivers.

Extending the number and the scope of IR matching operators, as well as making result aggregation formula customizable makes it difficult for end-users to control the scope of such queries. We plan to investigate processes supporting user needs description in a textual mode. A first stage may call the federated IR systems to extract criteria automatically (e.g., spatial and temporal entities and keywords) from a textual query. A second stage should ask the user to validate/adapt each criterion. Finally, a third stage may guide the user to specify expressiveness attached to each criterion. Indeed, our PIV\_spatial and PIV\_temporal IRSs process textual queries to extract spatial and temporal entities automatically (Gaio et al. 2008; Le Parc-Lacayrelle et al. 2007).

We plan to propose two approaches for the expressiveness specification stage. A first one, dedicated to regular users, would allow them to choose among predefined roles in order to add expressiveness to any criterion. A second one, dedicated to advanced users, may propose them to customize explicitly their criteria with modal ( $\wedge$ ,  $+$ ,  $-$ ) and matching ( $=$ ,  $\subseteq$ ,  $\cap$ ) operators.

Regarding the MIDR\_2010 test collection, a larger corpus and query collection need to be constituted for conducting further experiments.

Moreover, one can note that the proposed language completeness is not ensured although CMRP supports “or” and “and” operators via requirement and preference operators. However, this evaluation delivers some interesting preliminary findings and this contribution may stimulate the use of unexploited operators in GIRSs, as well as in other IRSs (e.g., Lucene (Gospodnetić and Hatcher 2005)).

## References

- Aula A, Khan RM, Guan Z (2010) How does search behavior change as search becomes more difficult? CHI'10, pp 35-44
- de Borda J-C (1781) Mémoire sur les élections au scrutin. Histoire de l'Académie Royale des Sciences
- Costa Pereira C, Dragoni M, Pasi G (2009) Multidimensional Relevance: A New Aggregation Criterion. ECIR'09, pp 264-275
- Croft B, Harabagiu S (2000) Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval. Kluwer Academic Publishers
- Farah M, Vanderpooten D (2008) An outranking approach for information retrieval. Inf Retr 11:315-334.
- Fernández M, Vallet D, Castells P (2006) Probabilistic Score Normalization for Rank Aggregation. ECIR'06. pp 553-556
- Fox EA, Shaw JA (1993) Combination of Multiple Searches. In: Harman DK (ed) TREC-1, pp 243-252
- Gaio M, Sallaberry C, Etcheverry P, et al. (2008) A global process to access documents' contents from a geographical point of view. Jour. of Vis. Lang. And Comp.. Academic Press, Inc., Orlando, FL, USA, pp 3-23
- Gan Q, Attenberg J, Markowetz A, Suel T (2008) Analysis of geographic queries in a search engine log. LocWeb'08, pp 49-56
- Gospodnetić O, Hatcher E (2005) Lucene in Action. Manning Publications
- Hubert G, Cabanac G, Sallaberry C, Palacio D (2011) Research and Advanced Technology for Digital Libraries. TPD'11, pp 118-129
- Jones R, Zhang WV, Rey B, et al. (2008) Geographic intention and modification in web search. Int. Jour. of Geo. Inf. Sci. 22:229-246.
- Kelly D, Fu X (2007) Eliciting better information need descriptions from users of information search systems. Inf. Process. Manag. 43:30-46.
- Larson RR (2009) Geographic Information Retrieval and Digital Libraries. ECDL'09, pp 461-464
- Liesaputra V, Witten IH, Bainbridge D (2009) Searching in a Book. ECDL'09, pp 442-446

- Lucas W, Topi H (2002) Form and function: the impact of query term and operator usage on Web search results. *J. Am. Soc. Inf. Sci. Technol.* 53:95-108.
- Manning CD, Raghavan P, Schütze H (2008) Introduction to Information Retrieval. 4,123,192,403.
- Martins B, Borbinha J, Pedrosa G, et al. (2007) Geographically-aware information retrieval for collections of digitized historical maps. *GIR'07*, pp 39-42
- Ounis I, Amati G, Plachouras V, et al. (2005) Terrier Information Retrieval Platform. *ECIR'05*, pp 517-519
- Palacio D, Cabanac G, Sallaberry C, Hubert G (2010) On the evaluation of Geographic Information Retrieval systems: Evaluation framework and case study. *Int. Jour. of Digit. Lib. (IJDL)* 91-109.
- Palacio D, Sallaberry C, Gaio M (2011) Generalize geographic information to combine IR results. *AGILE'11*, pp 151-153
- Le Parc-Lacayrelle A, Gaio M, Sallaberry C (2007) La composante temps dans l'information géographique textuelle. *Revue Doc. Num.* 10:129-148.
- Pfoser D, Efentakis A, Hadzilacos T, et al. (2009) Providing Universal Access to History Textbooks: A Modified GIS Case. *W2GIS*. pp 87-102
- Rasolofy Y, Hawking D, Savoy J (2003) Result merging strategies for a current news metasearcher. *Inf. Process. Manag.* 39:581-609.
- Sanderson M, Kohler J (2004) Analyzing Geographic Queries. *Proceedings of the Workshop on Geographic Information Retrieval, SIGIR*
- Silverstein C, Marais H, Henzinger M, Moricz M (1999) Analysis of a very large web search engine query log. *SIGIR Forum* 33:6-12.
- Spink A, Wolfram D, Jansen MJB, Saracevic T (2001) Searching the Web: the public and their queries. *J. Am. Soc. Inf. Sci. Technol.* 52:226-234.
- Tukey JW (1977) *Exploratory data analysis*. Addison-Wesley
- Usery EL (1996) A feature-based geographic information system model. *Photogrammetric Engineering & Remote Sensing* 62:833-838.
- Voorhees EM, Harman DK (1999) Overview of the Seventh Text REtrieval Conference (TREC-8). *TREC-8*. pp 1-23
- White RW, Morris D (2007) Investigating the querying and browsing behavior of advanced search engine users. *SIGIR'07*, pp 255-262
- Williamson DF, Parker RA, Kendrick JS (1989) The Box Plot: A Simple Visual Method to Interpret Data. *Ann Intern Med* 110:916-921.