

Contents lists available at ScienceDirect

Applied Soft Computing Journal



journal homepage: www.elsevier.com

Topic relevance and diversity in information retrieval from large datasets: A multi-objective evolutionary algorithm approach

Rocío L. Cecchini, Carlos M. Lorenzetti*, Ana G. Maguitman, Ignacio Ponzoni

Instituto de Ciencias e Ingeniería de la Computación (ICIC), CONICET, UNS, San Andrés 800, Campus Palihue, Bahía Blanca, Argentina

ARTICLE INFO

Article history: Received 30 April 2016 Received in revised form 24 September 2017 Accepted 10 November 2017 Available online xxx

Keywords: Query reformulation Information retrieval Topic modeling Diversity preservation

ABSTRACT

Enabling effective information search is an increasing problem, as technology enhances the ability to publish information rapidly, and large quantities of information are instantly available for retrieval. In this scenario, topical search is the process of searching for material that is relevant to a given topic. Multi-objective evolutionary algorithms have demonstrated great potential for addressing the topical search problem in very large datasets. In an evolutionary approach to topical search, a population of queries is automatically generated from a given topic, and the population of queries then evolves towards successively better candidate queries. Despite the promise of this approach, previous studies have revealed a common genotypic phenomenon: throughout evolution, the population tends to converge to almost identical sets of terms. This situation reduces the solution set to a few queries and leads to the exploration of a very limited region of the search space, which constitutes a limitation when users require different options from a topical search tool. This paper proposes and evaluates strategies to favor diversity in evolutionary topical search. These strategies rely on novel fitness functions, different parameterization for the crossover and mutation rates, and the use of multiple populations to favor diversity preservation. Experimental results conducted using these strategies in combination with the NSGA-II algorithm on a dataset consisting of more than 350,000 labeled web pages indicate that the proposed strategies show great promise for searching very large datasets, by helping to achieve query and search result diversity without giving up precision.

© 2017.

1. Introduction

Effective information retrieval from very large datasets has become increasingly important as technology enhances the ability to publish information rapidly and large quantities of information are instantly available for retrieval. In this scenario, topical search emerges as a useful procedure that allows seeking material related to a topic of interest [1,2].

There are many Big Data applications in which topical search has proven to be useful. For instance, topical search can be applied to support task-based information search, where an intelligent assistant monitors the user's activity and retrieves information by contextualizing user's information requirements [3,4]. Another application in which topical search can be usefully applied is deep Web access, where queries are formulated to harvest deep Web resources [5]. Topical search can also be applied to satisfy persistent information interests, where a software agent can play an intermediary role between a search interface and a user. In this kind of applications, the agent learns the user's interest profile and uses this profile to construct queries automatically and to filter results, offering information that of interest the is to

* Corresponding author.

user over an extended period of time [6-8]. Finally, topical search shows great promise in the area of *opinion mining*, where topical search is applied in social media to collect opinions about specific topics [9,10].

Several techniques exist to infer and represent a topic of interest to a user. This can be done by monitoring the user's activities, such as web pages or social web content being recently visited, documents being edited or emails being written [11-14]. Once the topic of interest has been represented, a new major challenge is to develop mechanisms to search for material related to the given topic.

By designing computational strategies for generating topical queries it becomes possible to automate the process of retrieving topic-relevant content. The development of such strategies will highly depend on assessing the effectiveness of the generated queries. This assessment is difficult because multiple objectives need to be considered to guide the query formulation process. In a broad sense, a query is effective if it retrieves the material the user is looking for. Given a predefined set of relevant items for a topic of interest, classical performance metrics such as precision and recall can be computed for individual queries to evaluate topical relevance. The precision of a query is the fraction of documents retrieved (when posing the query) that are relevant, while the recall of a query is the fraction of relevant documents that are actually retrieved.

Email addresses: rlc@cs.uns.edu.ar (R.L. Cecchini); cml@cs.uns.edu.ar (C.M. Lorenzetti); agm@cs.uns.edu.ar (A.G. Maguitman); ip@cs.uns.edu.ar (I. Ponzoni)

As discussed in recent studies, evolutionary algorithms (EAs) have proven to be successful in dealing with Big Data problems [15,16]. More specifically, the problem of topical search can be seen as a multi-objective optimization problem where the objective function to be maximized quantifies the effectiveness of a query [17,18].

In the topical-search optimization problem the initial set of candidate solutions is defined as the set of possible queries that can be presented to a search interface. Another aspect of this optimization problem is that the query space is a high-dimensional space, where each possible term in a query accounts for a new dimension. This kind of problems are computationally expensive and cannot be effectively solved using analytical methods. Besides, the problem of query optimization does not have optimal substructure, which means that an optimal solution cannot be constructed efficiently from optimal solutions to its subproblems [19]. Therefore, existing methods to solve complex problems by breaking them down into simpler steps are not effective for our purpose. On the other hand, a query can be considered effective even if it is not an optimal one, at the same time as multiple and diverse queries can provide satisfactory results. Therefore, we may be interested in finding many near-optimal queries rather than a single optimal one.

On the basis of the characteristics described above, EAs [20,21] are applicable to the problem of learning to automatically formulate optimal (or near-optimal) topical queries. Also, as mentioned above, more than one, possibly conflicting objectives can be defined to measure query effectiveness, resulting in a Multi-Objective Optimization Problem (MOOP) that may need to balance different criteria, such as precision, recall or other metrics.

In previous research work we have successfully applied Multi-Objective Evolutionary Algorithms (MOEAs) to evolve a population of topical queries [17,18], where the objectives to be maximized were precision and recall. However, evaluating individual queries in terms of precision and recall only provides a partial solution to the problem of measuring the performance of topical query generation. For the topical-search applications discussed earlier, additional desiderata for query generation strategies must be considered. In these applications, *the generated queries should be evaluated collectively rather than independently of each other*. Therefore, besides attempting to achieve high performance for the individual queries, *it is important to jointly attain high coverage and diversity while preserving local coherence*. In other words, we expect each topical query to be specific enough to return mostly relevant results, but the whole set of queries needs to be diverse enough not to miss many relevant results.

The problem of diversification of search results has been broadly explored by the information retrieval community (see [22] for a recent survey). In the meantime, population diversity has long been considered a critical issue in the performance of EAs [23]. However, to the best of the authors' knowledge, the solutions coming from the fields of information retrieval and evolutionary computation have been treated separately. Given the suitability of EAs to deal with the problem of topical search and the importance of diversification of search results, it is of great importance to investigate diversity preservation strategies in the context of evolutionary topical search. This article reviews and evaluates different aspects associated with the problem of diversity preservation in MOEAs for the specific scenario of topical search.

The main contributions of this work include:

• The definition of specialized fitness functions, where query performance not only depends on the results retrieved by the individual query but also depends on the results returned by other queries in the same population. The new metrics involve a reformulation of the classical precision and recall metrics, and we refer to them as *retrospective precision* and *retrospective recall*. To help achieve diversity, the new metrics penalize queries that retrieve results that have been previously obtained by other queries in the same population.

- The use of variable mutation and crossover rates by applying *hypo-crossover* (low crossover probability), *super-mutation* (high mutation probability), and *hyper-mutation* (very high mutation probability).
- The use of multiple populations of queries, to overcome the problem faced when a single population of queries converges to almost identical sets of terms.

The proposed strategies are evaluated using classical performance metrics as well as ad-hoc ones specifically aimed at assessing the diversity of the generated collection of queries as well as the coverage of relevant results retrieved by the entire population of queries. Population diversity is evaluated at the genotypic level, by assessing the diversity among the queries themselves, and at the phenotypic level using the Pareto front of the solutions. Finally, by analyzing the global recall reached by the population as a whole, it is possible to evaluate diversity at another phenotypic level.

2. Background and related work

2.1. Topical search

The process of searching for online data can be guided by diverse objectives. There are essential differences between searching for information to satisfy a particular consultation need and searching for resources to support the process of topical search. Usually, the purpose of a consultation is to find an answer to a particular question quickly and accurately. On the other hand, the purpose of topical search is to seek material based on a topic of interest. Different from the task of fulfilling a consultation need, topical search usually does not require a high-speed response. On the other hand, while accuracy is important for topical search, additional criteria should be considered such as high coverage and variety of results.

This article focuses on designing methods for topical search that can be used for applications where high speed is not crucial, and the main focus is to achieve high recall, novelty, and diversity without giving up precision. These applications can be built on top of existing search interfaces and can be used in different scenarios, such as searching for material to augment knowledge models [3,24], fulfilling long-term information needs [6–8,4], collecting resources for topical Web portals [25], accessing the deep Web [5], systematically reviewing medical digital repositories or records to identify research literature or previous experiences relevant to a given disease [26], among others.

The design of topic-based search methods usually requires the definition of methods for topic extraction and modeling, query generation, query refinement, and topic-based filtering. A great variety of methods have been proposed to search for material based on a topic or thematic context [11,27–29,18,30]. Typically, these methods represent the user thematic context as a set of terms and automatically generate queries based on this representation.

A seminal context-based search system is the *Remembrance Agent* [31], which operates inside the Emacs text editor and continuously monitors the user's work to find relevant text documents, notes, and emails previously indexed. Another influential context-based assistant is the *Watson* system [11]. *Watson* uses contextual information from documents that users are manipulating to automatically generate Web queries from the documents, using a variety of term-extraction and weighting techniques for selecting suitable query terms, filtering the matching results and clustering similar HTML pages. In [28] three general approaches for searching in context are discussed: a query-rewriting technique, a rank-biasing technique, and an iterative filtering

techniques based on the use of a meta-search engine. An incremental method for query refinement based on the notions of topic descriptors and discriminators is introduced in [32]. Later [29], enhanced this method by endowing the system with the ability to learn new context-specific vocabulary. Another system called *Leibniz* [30] introduces the notion of "contextual insights" to provide users with information that is contextually relevant to the content that they are consuming or authoring. Some key aspects of the *Leibniz* system include the prediction of the user's focus of attention, the addition of context terms to the query and the adaptation of the results via post-processing.

The application of EAs for searching has mostly focused on the development of term-weight reinforcement techniques for query optimization (e.g., [33,34]). EAs have also proved to be useful for developing inductive query-by-example (IQBE) techniques with the ability to identify new documents that are similar to documents initially provided by users [35,36]. These techniques were shown to outperform traditional techniques based on relevance feedback. A comparative study of different MOEA strategies to learn Boolean queries in the context of the IQBE paradigm is presented in [37]. Other search approaches based on EAs have focused on refining the initial set of results using improved queries [38–40,17,18]. However, different from the current study, these approaches do not address the problem of diversity preservation.

2.2. Diversification of search results

The problem of diversification of search results has been extensively studied in the information retrieval community. A study presented in [41] discusses the importance of optimizing the relevance of results at the document set level instead of at the single document level. This is closely related to our goal of attaining coverage and diversity at the global level. However, instead of proposing strategies for attaining diversity, their goal is to compare clustering and diversification of search results using a unified evaluation framework, where the notion of *subtopic* plays a key role. A seminal work by [42] proposes to re-rank the list of results to reduce redundancy. This is achieved by using a metric called "marginal relevance" that is a linear combination of relevance and novelty. A document has a high score if it is both relevant to the query and contains minimal similarity to the previously retrieved results. More recently, a large number of frameworks for Web search result diversification that explicitly account for the various aspects associated with an underspecified query have been proposed (e.g., [43–45]). However, these approaches focus on methods aimed at identifying diverse results associated with a single query rather than on generating diverse queries associated with a topic of interest, as is done by the methods proposed in this article.

Several strategies for diversifying search results rely on query diversification mechanisms. A technique for suggesting alternative queries to Web users is proposed in [46]. This technique takes an initial query and suggests diverse queries that are semantically related to the original query. It relies on the use of query-URL click through data and is based on Markov random walk. Other techniques for diversifying query suggestions take advantage of existing information sources, such as topics from Wikipedia (e.g., [47]) or employ the WordNet ontology for identifying different senses of the original query (e.g., [48]). Existing methods for query diversification sometimes require an initial description defining a topic of interest that allows identifying diverse aspects (or subtopic). In this case, the goal is to generate diverse queries that can cover multiple aspects associated with a topic or task. In [49] query diversification is attained by taking a document from which multiple aspects can be identified. Since documents are typically long, they can be used generate several queries covering their varito

ous aspects or subtopics. These queries are then filtered by selecting those ones that are more relevant to the query document, as well as novel relative to the current query suggestion list. A comprehensive survey of search result diversification approaches is presented in [22]. It is worth mentioning that all of these proposals address the problem of diversification of search results by applying strategies considerably different from those presented in this article. In particular, none of these strategies rely on the application of EAs for topical retrieval.

2.3. Multi-objective evolutionary algorithms

EAs [20,21] are robust optimization techniques based on the principle of "natural selection and survival of the fittest," which claims "in each generation the stronger individual survives and the weaker dies" [50]. Therefore, each new generation would contain stronger (fitter) individuals than their ancestors.

To use EAs in optimization problems we need to define candidate solutions (also called individuals) by chromosomes consisting of genes and a fitness function to be minimized (or maximized). A population of individuals is maintained. The goal is to obtain better solutions after some generations. To produce a new generation, EAs typically use selection together with the genetic operators of crossover and mutation. Parents are selected to produce offspring, favoring those parents with highest values of the fitness function. Crossover of population members takes place by exchanging subparts of the parent chromosomes (roughly minicking a mating process), while mutation is the result of a random perturbation of the chromosome (e.g., replacing the value of a gene by another allele).

Although crossover and mutation can be implemented in many different ways, their fundamental purpose is to explore the search space of candidate solutions (also known as *decision variable space*), improving the population at each generation by adding better offsprings and removing inferior ones. It is important to remark that this exploration is guided by the selection operator, which uses the fitness values associated with the candidate solutions. Therefore, these fitness values constitute the space where the selection works (also known as *objective function space*). For this reason, the fitness function defines a mapping between the *decision variable space* and the *objective function space*. Several introductory books and survey articles are available for a complete study of the topic [51–53,21].

In multi-objective optimization problems (MOOPs) the quality of a solution is defined by its performance in relation to several, possibly conflicting, objectives. Traditional methods are very limited because, in general, they become too expensive as the size of the problem grows [54,55]. In this context, MOEAs became suitable techniques based on EAs for dealing with MOOPs [56,57,21]. There are many approaches to multi-objective optimization using MOEAs and, in general, they can be classified in Pareto or non-Pareto EAs. In the first case, the evaluation is made following the Pareto dominance concept [58].

Dominance is a partial order that could be established among vectors defined over an *n*-dimensional space. Using a multi-objective fitness function it is possible to define a relation between vectors x_i in the *decision variable space* and vectors u_i in the *objective function space*. Therefore, a Pareto optimal solution is an individual, x_j , within the *decision variable space*, whose corresponding vector components in the *objective function space*, u_j , cannot be all simultaneously improved for any other u_i , with $i \neq j$. A non-dominated set of a feasible region in the *objective function space* defines a Pareto Front (PF), and the set of its associated vectors in the *decision variable space* is called the Pareto Optimal Set. The Pareto-based algorithms use the concept of domination for the selection mechanism to move a population toward the Pareto Front. In contrast, in the non-Pareto EAs, the objectives are combined to obtain a single evaluation value to be used for the selection mechanism.

The results discussed in this work are based on the application of NSGA-II (Non-dominated Sorting Genetic Algorithm – II) [59], which is one of the most studied and efficient Pareto EAs and is based on non-domination sorting. We refer the reader to [59,56] for a detailed explanation of NSGA-II and other MOEAs. The PISA platform [60] was used to implement the strategies analyzed in this work.

2.4. Diversity preservation in evolutionary algorithms

There are two central goals in Pareto-based MOOPs: (1) to discover solutions as close to the Pareto-optimal solutions as possible, and (2) to find solutions as diverse as possible in the obtained non-dominated front. The first goal requires a search towards the Pareto-optimal region, while the second goal requires a search along the Pareto-optimal front. While the first purpose could be partially guided by the evolutionary process, the second one demands a selection mechanism that considers diversity.

According to traditional assessments of diversity [61,62], a diverse set of solutions should cover the entire Pareto-optimal region uniformly. This means there are two different aspects of diversity: the extent or spread of extreme solutions and the distribution or relative distance among solutions. Alternative approaches could be adopted to maintain diversity. Typically, a measure diversity among individuals needs to be computed, which can be based on either the individuals' genotype (i.e., computed in the *decision variable space*) or phenotype (i.e., computed in the *objective function space*).

A widely adopted measure to assess diversity is *Max-Sum Diversity* [63], which is used to maximize the sum of distances among the individuals. Alternatively, the Max-Min Diversity measure [64] is used to maximize the minimum distance among the individuals. Other measures include the Min-P-Center Diversity measure [65], in which the greatest of the minimum distances is minimized and the Max-Mean Diversity measure [66], which is aimed at maximizing the average distance among the individuals. In [67] a MOEA is proposed to simultaneously optimize several diversity measures.

Our problem domain is suitable for computing diversity by considering either the genotype or phenotype representation of the individuals. For example, genotypic diversity can be promoted by considering the vector representations of the individual queries, computing the similarity among them and penalizing those queries that are similar to many other queries. On the other hand, an approach based on phenotypic similarity could take into consideration information associated with the retrieval results of each query. A simple implementation of query phenotypic similarity could consist in counting the number of common results returned by two queries.

Several Niching methods have been developed to address issues related to diversity. The best-known strategies aimed at niching are crowding [68] and fitness sharing [69]. In both cases, the goal is to obtain solutions uniformly distributed over the PF. Crowding methods favor individuals from less crowded regions. A notion of crowding distance is defined to quantify how crowded a region is. The crowding distance of a solution x_i is a measure of the *decision variable space* around x_i which is not occupied by any other solution in the population. Details on how this measure is implemented can be found in [57].

Fitness sharing is a diversity-preserving mechanism that modifies the landscape by reducing the payoff in densely populated regions. The fitness of an individual is lowered by an amount nearly equal to the number of similar individuals in the population. The main disadvantage of this method is that it is necessary to provide the niche size parameter, which defines a neighborhood of solutions in the *objective function space*.

A key issue with niching techniques has to do with the computational cost associated with the estimation of some measure of similarity among individuals. An appealing and less expensive technique that favors the exploration of diverse solutions is local selection [70]. This approach is similar to but more efficient than fitness sharing. It has the advantage of allowing parallel implementation for distributed tasks, an important pro in the information retrieval scenario. This is achieved by evaluating the fitness function by an external environment that provides appropriate data structures for maintaining shared resources. This technique is amenable to the information retrieval task, where each retrieved document can be marked so that the same document does not yield payoff multiple times.

Beyond the study of the diversity of the individuals at the traditional genotypic and phenotypic levels, a population fitness measure is introduced in this work to evaluate the diversity of a population as a whole. The concept of population fitness is present in biology [71], supported by the principle of *community heritability* attributed to Goodnight [72,73]. This principle said: "if the interactions among the members of the community are passed intact from the *parent* community to the *offspring* community, the interaction will be heritable at the community level". Therefore, if the population is considered as the unit of survival, a population with high fitness is one that has a high tendency to out-reproduce other populations. Note that population fitness can be unrelated to the concept of average fitness in the population.

In EAs, an example of population fitness is incorporated in the Parisian Approach [74], which is denoted as *global fitness*. In this algorithm, a global solution is built by the combination of the information provided by several members of the population, whose individual performances are evaluated by a *local fitness*. In our application, we use a population fitness measure named *global recall*, which in turn gives rise to another population fitness measure named *global f-measure*. The *global recall* function computes the recall over the complete set of queries contained by the whole population. In the meantime, the *global f-measure* function combines recall at the global (population) level and precision at local (individual) level.

Another strategy usually exploited for diversity improvement is based on the concept of multiple populations [75]. The idea behind these EAs consists in defining subpopulations which evolve independently; thus the unique features of each subpopulation can be effectively preserved, and the diversity of the entire population is benefited. In some algorithms, these subpopulations can exchange information, as it is the case in *Multi-population Coevolutionary Algorithms* [76], but in other cases, a complete isolated evolution is executed over each subpopulation [77]. In this work, this last alternative of multi-population algorithm is explored.

Fig. 1 presents a summary of the background concepts and related work discussed in this section.

3. A MOEA framework for diversity-preserving topical search

This section reviews the general framework to evolve topical queries presented in [18]. It also describes how the original framework was adapted and augmented to cope with the problem of diversity preservation. In [18], the authors propose to apply two MOEAs to the problem of retrieving documents based on a given topic. The first MOEA uses NSGA-II to evolve a population of queries guided by two objectives: (1) to attain high query precision, and (2) to attain high query recall. The second MOEA is a non-linear aggregative algorithm that is also driven by precision and recall. Also, three non-evolutionary query generation methods were tested to determine if the

Applied Soft Computing Journal xxx (2017) xxx-xxx



MOEAs show improvements over the non-evolutionary approaches. The non-evolutionary methods include the following:

- *Baseline*: the queries are generated by taking words from the description associated with each topic.
- Divergence from Randomness mechanism with Bose–Einstein statistics: this method, also known as Bo1 [78], is considered one of the most highly effective successors of the classical Rocchio's query-refinement method [79].
- Supervised Divergence from Randomness mechanism with Bose–Einstein statistics: this method, to which we refer to as Bo1*, is a supervised version of Bo1 and hence it is considered even more effective because it uses relevance information to determine the most informative terms to be used for query expansion.

The evaluation of these methods over more than a hundred topics reported in [18] showed that the queries evolved by the two MOEAs attain higher precision and recall than the queries generated by applying the non-evolutionary techniques. As a consequence, this work will not compare the performance of MOEAs for query generation against non-evolutionary methods. Instead, it will focus on the evaluation of different computational strategies that can be applied to MOEAs to achieve high performance in terms of precision and recall, while attempting to attain diversity. To achieve this, we construct several computational strategies based on the combination of different fitness functions and different mutation and crossover rates. Each strategy uses the NSGA-II algorithm to evolve topical queries starting with a population of queries composed of terms extracted from an initial description of the given topic. The topic description is available in the ODP distribution files used in our experiments (as explained in Section 4.1) and typically consists of a few sentences briefly describing the content of the pages indexed under that topic. It is important to note that the topic description is a good starting point for the query generation process since it is a summary provided by an editor who is familiar with the topic.

As generations pass, queries associated with improved search results will predominate. Furthermore, the mating process continually combines these queries in new ways, generating ever more sophisticated solutions. In particular, the mutation mechanisms can be implemented in such a way that novel terms, i.e., terms that are not in the initial user context, are brought into play.

3.1. Population and representation of chromosomes

The decision variable space Q contains all possible queries that can be formulated to a search interface. Thus the population of chromosomes is a multiset of queries (note that repetitions are possible). In other words, each population is an element of \mathbb{N}^Q . Each chromosome is represented as a list of terms $t^1, t^2, ..., t^q$, where each term t^i corresponds to a gene that can be manipulated by the genetic operators. The population is initialized with a fixed number of queries randomly generated with terms from the topic description. Note that while initial terms are randomly chosen to form queries, the selection mechanism of the implemented evolutionary strategies will favor the most effective terms in subsequent generations. The number of terms in each of the initial queries will be random, with a constant upper bound on the query size. Although the sizes of the initial queries are never more than a predefined constant, the sizes of some queries in subsequent generations can exceed this limit. This is because applying the crossover operator can change the offspring size.

The terms forming each query are not sorted and may be repeated within a query. While term repetition does not have an impact on matching for retrieval, it may have an impact on document ranking, and hence it can affect some fitness functions such as Precision@10 (described in the next section). While all terms in the initial population of queries come from the initial topic description, novel terms can be included in the queries after mutation takes place. These novel terms are obtained from a *mutation pool* (described in Section 3.4), which is an ever increasing set of terms that may or may not be part of the initial context.

In our analysis, we have looked at the evolution of a single population and also at the parallel evolution of multiple independent populations, as a strategy for addressing the problem of diversity preservation.

3.2. Fitness functions

A fitness function defines the criterion for assessing the quality of a query. Assuming that it is possible to distinguish relevant documents from non-relevant ones, it is possible to measure the precision of a query as the fraction of retrieved documents that are relevant. Moreover, if the set of relevant documents is known in advance, it is possible to measure the recall of a query as the fraction of relevant documents that have been retrieved.

In the adopted model, queries are interpreted following the disjunctive semantics, which means that a single matching term is sufficient to retrieve a document. Let D_t be the set containing all the documents associated with topic t, including those in its subtopics. To define *precision* we associate with the search space Q and topics T a function *Precision* : $Q \times T \rightarrow [0, 1]$ that can numerically evaluate an individual query q for a given topic t as follows:

$$\operatorname{Precision}(q,t) = \frac{|A_q \cap D_t|}{|A_q|},$$

where A_q is the answer set associated with q, containing all the documents returned by a search engine when q is used as a query.

The selection of a weighting model for assigning scores to the retrieved documents will strongly influence query performance. For this purpose, we have used a vector representation of the query. Documents are weighted based on the TF-IDF scheme (term frequency-inverse document frequency) [80] and similarity between queries and documents is computed based on the cosine similarity metric [81]. Therefore, a ranking over the retrieved documents is possible. This retrieval configuration will typically result in a large number of matches, sorted by their similarity to the query vector. Therefore, rather than looking at precision, we take *precision at rank 10* [82], which is the fraction of the top 10 retrieved documents which are known to be relevant. To define *precision at rank 10* we associate with the search space *Q* and topics *T* a function *Precision* @ $10: Q \times T \rightarrow [0, 1]$ that can numerically evaluate an individual query *q* for a given topic *t* as follows:

Precision@10(q,t) =
$$\frac{|A_q^{10} \cap D_t|}{|A_q^{10}|}$$
,

where D_t was defined above and the set A_q^{10} consists of the top-10 ranked documents in the answer set A_q returned by a search engine when q is used as a query. Notice that the size of A_q^{10} might be less than 10 if the search engine returns less than 10 results for the query q.

On the other hand, *recall* is defined as a function *Recall* : $Q \times T \rightarrow [0, 1]$ as follows:

$$\operatorname{Recall}(q,t) = \frac{|A_q \cap D_t|}{|D_t|},$$

where A_a and D_t were defined above.

We define two novel fitness functions to penalize those queries that do not help to achieve diversity, namely *retrospective precision at rank 10* and *retrospective recall*, which are analogous to the classical *Precision@10* and *Recall* metrics. The approach adopted to define these fitness functions is bio-inspired and mimics a multiset of individuals (queries) in a population competing to consume a finite resource (relevant documents). Therefore, the outputs of the retrospective fitness functions for a given individual q_i , not only depend on q_i but also depend on other individuals in the same population. Let *P* be a population of queries. To formally define the retrospective fitness functions it is necessary to assume a total order \prec on the queries that constitute *P*. Given the totally ordered multiset *P* and a query $q_i \in P$, we define the set R_{P,q_i}^{10} as the union of the top ten results returned by each of the queries that precedes q_i :

$$R_{P,q_i}^{10} = \bigcup_{q_j \prec q_i} A_{q_j}^{10}.$$

The set R_{P,q_i}^{10} can be thought of as the set of resources that have been consumed by other individuals preceding q_i , assuming each individual can consume up to ten resources. Then the set R_{P,q_i}^{10} is used to numerically evaluate an individual query q_i in terms of *retrospective precision at rank 10*, using the function *Retro-Precision@10*: $\mathbb{N}^Q \times Q \times T \rightarrow [0, 1]$ defined as follows:

Retro-Precision@10(P,q_i,t) = $\frac{(|A_{q_i}^{10} \cap D_t) \setminus R_{P,q_i}^{10}|}{|A_{q_i}^{10} \setminus R_{P,q_i}^{10}|}.$

In a similar way, let the set R_{P,q_i} be the union of the results returned by each of the queries q_1, q_2, \dots, q_{i-1} , defined as:

$$R_{P,q_i} = \bigcup_{q_j \prec q_i} A_{q_j}.$$

The set R_{P,q_i} represents the set of resources that other individuals preceding q_i have already consumed, without imposing a limit on the number of resources that each individual could consume. To compute *retrospective recall* we define the function *Retro-Recall*: $\mathbb{N}^Q \times Q \times T \rightarrow [0, 1]$ as follows:

$$\text{Retro-Recall}(P, q_i, t) = \frac{|(A_{q_i} \cap D_t) \setminus R_{P,q_i}|}{|D_t \setminus R_{P,q_i}|}$$

These fitness functions are aimed at evaluating the *precision at rank 10* and *recall* of a query in such a way that those results that have been retrieved by a previous query (based on the established ordering) in the same population are disregarded. Therefore, the established ordering has an important impact on the values of these fitness functions, since it determines how the results retrieved by a query affect the fitness of those queries that are evaluated subsequently.

In this work, the ordering is defined in such a way that the best performing queries are given priority. If two queries are equally effective, then a random choice is made to favor one over the other. This priority imposed over individuals attempts to mimic the natural fact that the fittest individuals in a population tend to have an advantage over the weakest individuals when competing for a finite resource. Therefore, for each generation it is necessary to rate each query based on its performance. To avoid circularity and to obtain a single measure of query performance we take the harmonic mean of the classical (non-retrospective) *Precision* and *Recall* functions, resulting in an aggregated function $F - measure : Q \times T \rightarrow [0, 1]$ defined as follows:

$$F - \text{measure}(q, t) = 2 \cdot \frac{\text{Precision}(q, t) \cdot \text{Recall}(q, t)}{\text{Precision}(q, t) + \text{Recall}(q, t)}$$

The *F-measure* function returns a performance rating for each query, allowing to establish a total order for the population of queries at each generation. If two or more queries have the same *F-measure* value, then we force a total order by arbitrarily breaking any existing tie.

3.3. Genetic operators

A new generation in our EAs is determined by a set of operators that select, recombine and mutate queries of the current population.

- Selection: After ranking the solutions, a population of *n* offsprings is created using recombination, mutation and a diversity-preserving binary tournament selection operator based on crowding distances. This operator, known as *crowded tournament selection* is in charge of the NSGA-II selection procedure. According to it, a solution x_i wins a tournament with another solution x_j if the following conditions are true:
 - If solution x_i has a better (smaller) rank than solution x_i ,
 - If the ranks of both solutions are the same, the solution located in the less crowded region is preferred.
- **Crossover:** Some of the selected queries are carried out into the next generations as they are, while others are recombined to create new queries. The recombination of a pair of parent queries into a pair of offspring queries is carried out by copying selected terms from each parent into the descendants. The crossover operator used in our proposal is known as single-point. It results in new queries in which the first *n* terms are contributed by one parent and the remaining terms by the second parent, where the crossover point *n* is chosen at random. Our analysis will consider the application of crossover at different probability rates. In particular, we will study the use of crossover at a typical probability rate (P = 0.7), and *hypo-crossover*, which refers to a low crossover probability (P = 0.35).
- **Mutation:** Small random changes can be produced to the new population of queries. These changes consist in replacing a randomly selected query term t^q by another term t^p . The term t^p is obtained from a *mutation pool* (described next). Besides testing mutation at a typical probability rate (P = 0.03), we will analyze the effect of applying *super-mutation* (P = 0.3) and *hyper-mutation* (P = 0.7).

3.4. Mutation pool

The mutation pool is a set that initially contains terms extracted from the topic description under analysis. As the system collects relevant content, the mutation pool is updated with all the non-stopword terms found in the relevant documents that have been recovered. This procedure brings new terms to the scene, allowing a broader exploration of the search space and making it possible to learn new vocabulary from relevant documents.

4. Evaluating the diversity-preserving strategies

4.1. Data collection and experimental setup

To run our evaluations we used the evaluation framework described in [83]. This framework was built by collecting the URLs associated with 448 topics from the Open Directory Project (ODP – http: //dmoz.org). The topics were selected from the third level of the ODP hierarchy. Some constraints were imposed on this selection with the purpose of ensuring the quality of our corpus. For each topic, we collected all of its URLs as well as those in its subtopics. The minimum size for each selected topic was 100 URLs and the language was restricted to English. The total number of collected pages was more than 350,000. The Terrier framework [84] was used to index these pages and to run our experiments. We used the stopword list provided by Terrier and Porter stemming was performed on all terms. We divided each topic in such a way that two-thirds of its pages were used to create a training index and the remaining one-third of the corpus was used for testing.

The evaluation was carried out by randomly choosing two different topics out of the 448 indexed topics. The topics selected to carry out the evaluations were BODY PAINTING and BIOINFORMATICS. For each of these two topics, we used its ODP description to characterize it as an initial set of terms. The ODP topic description is available in the ODP distribution files and typically consists of a few sentences briefly describing the content of the pages indexed under that category.

A statistical study was conducted to determine significant differences between the performance of several combinations of strategies described in the following paragraphs. The study consisted on the calculation of mean values (over average population values) across all runs and their corresponding confidence intervals at the 95% level. The statistical evaluations were performed in two stages. In the first stage we used the training set to evolve queries and monitored the performance (using average population values) of the queries at each generation. For each of the two topics used in the evaluation, 50 runs of the MOEAs were executed. The number of generations for each run was set to 166 while the population consisted of 100 queries. These values were strategically selected after a previous analysis that allowed us to conclude that increasing the number of generations or the size of the population beyond these values did not have a significant effect on the results.

For each analyzed topic the population of queries was randomly initialized using its ODP description. The size of each query was a random number between 1 and 32. We should point out that existing search engines use up to a fixed number of query terms and ignore subsequent ones (e.g., Google's query size limit is 32 terms). Applying the crossover operation could eventually increase the query size beyond this limit. The terms that go beyond that limit are ignored when the query is formulated (because they occur beyond the query size limit), but they may be taken into account in subsequent generations when these terms become part of an offspring query after recombination takes place. For the second stage we took the queries evolved during the first stage and evaluated them on the test set. The goal was to determine if the evolved queries for a particular topic were effective on a new corpus (the test set). Note that the training and test sets contain the same topics (and therefore the same topic descriptions) but different documents.

With the purpose of comparing different strategies, we first evaluated MOEAs that employed various combinations of fitness functions. Each combination attempted to optimize two different objectives, one aimed at attaining high precision (e.g., *Precision, Precision@10* or *Retro-Precision@10*), and the other aimed at attaining high recall (e.g., *Recall* or *Retro-Recall*). Finally, we carefully selected three different combinations of fitness functions to report a manageable, yet representative subset of results:

| Co1 | Precision@10 and Recall. |
|-------|--------------------------------------|
| Co2 | Precision@10 and Retro-Recall. |
| Co3 | Retro-Precision@10 and Retro-Recall. |
| C 1.3 | |

Combination **Co1** involved classical measures of performance and therefore was taken as a baseline for comparison purposes. Also, **Co1** was used in conjunction with other strategies, such as different crossover and mutation rates, as well as multiple populations to analyze the independent effect of these additional strategies. Combinations **Co2** and **Co3** were used to analyze the effect of incorporating the two retrospective fitness functions (i.e., *Retro-Recall* and *Retro-Precision@10*), one at a time. In the same way as for **Co1**, these two combinations were analyzed in conjunction with different crossover and mutation rates, as well as with single and multiple populations. This allowed to analyze the cumulative effect of combining more than one diversity preservation strategies.

The settings for the mutation and crossover rates that we report in this article are the following:

| Normal | The probability of crossover was set to 0.7 (normal) and the probability of mutation was set to 0.03 (normal). | | | | |
|-------------|--|--|--|--|--|
| НуроС | The probability of crossover was set to 0.35 | | | | |
| | (hypo-crossover) and the probability of mutation was set to 0.03 (normal). | | | | |
| SuperM | The probability of crossover was set to 0.7 (normal) | | | | |
| | and the probability of mutation was set to 0.3 (super-mutation). | | | | |
| HyperM | The probability of crossover was set to 0.7 (normal) | | | | |
| | and the probability of mutation was set to 0.7 (hy- | | | | |
| | per-mutation). | | | | |
| HypoC+Super | M The probability of crossover was set to 0.35 | | | | |
| | (hypo-crossover) and the probability of mutation | | | | |

(hypo-crossover) and the probability of mutation was set to 0.3 (super-mutation).

These rates were selected based on the expertise obtained in previous work [see Ref. 17]. The analysis was carried out for both single and multiple populations. In our setting, a single population consists of a multiset of queries that interact with each other. The interaction between individual queries in the same population is given in the usual way for EAs: through competition (as the result of applying binary tournament selection) and through recombination (as the result of applying the crossover operator to two queries). It is worth noting that when the retrospective fitness functions are applied (i.e., Retro-Precision@10 or Retro-Recall), individual queries also compete at a different level, with a bias in favor of those queries that are evaluated first. On the other hand, for the multi-population case, different populations of queries evolve in such a way that there is no interaction between any two populations. In our evaluations, analyzed we have single-population MOEAs with 100 individuals and multi-population MOEAs consisting of 10 independent populations with 100 individuals each.

Our analysis includes the results of evaluating the single- and multi-population strategies in conjunction with the three different combinations of fitness functions (Co1, Co2 and Co3), and, in turn, combined with the five different settings for mutation and crossover rates (Normal, HypoC, SuperM, HyperM, and HypoC+SuperM). As a result, a total number of 30 different computational strategies are reported for each of the analyzed topics. The comparison is based on results averaged over 100 queries and 50 runs.

As mentioned earlier, the computational strategies proposed had two central goals: (1) to discover optimal or near-optimal solutions with respect to the given performance measures, and (2) to find solutions as diverse as possible. In the following sections we analyze the degree to which these goals are met by the computational strategies introduced above. The application of multi-population strategies as an upgrade to the single population ones only has an impact on diversity, without affecting the performance metrics of Precision@10 and Re*call*. This claim was empirically tested and naturally follows from the approach adopted to implement the multi-population strategies analyzed in this work, where each of the populations of queries evolves as a single population, without any interaction with the other populations. Therefore, in Section 4.2 we report the Precision@10 and Recall performance values achieved by the 15 single-population strategies. In addition, we present a visual analysis of the Pareto Fronts achieved by these strategies. This is followed by a study of different aspects of diversity, extending our analysis to both single and the multiple populations.

The analysis of population diversity is reported in Section 4.3, where we initially focus on the similarity between queries using the well-known cosine similarity formula [81]:

Cosine–Similarity
$$(q_1, q_2) = \frac{\vec{q}_1 \cdot \vec{q}_2}{\|\vec{q}_1\| \cdot \|\vec{q}_2\|}$$

where (q_1, q_2) is a pair of queries and \vec{q}_i is the TF vector representation of q_i , based on the terms in the query. A global similarity coefficient value is then computed by averaging the similarity between all pairs of queries (q_i, q_j) in the population of queries $P = \{q_1, q_2, ..., q_n\}$:

$$\begin{aligned} \text{Similarity}(P) \\ &= \frac{\sum_{q_i, q_j \in P, i \neq j} \text{Cosine} - \text{Similarity}(q_i, q_j)}{|P| \cdot (|P| - 1)}. \end{aligned}$$

To provide additional insights on population diversity, the distributions of the Pareto Fronts for multiple populations of queries is reported and contrasted with those obtained for single populations.

Section 4.4 provides an analysis of the population performance at a global level by reporting classical global performance metrics as well as ad hoc ones. A standard method to quantify population performance is by computing the area of a Pareto-approximated front in objective space. This area can be calculated in a similar way to the area under the receiver operating characteristic curve (or AUC), using the trapezoidal technique [56]. AUC values range between 0 and 1, with maximum accuracy achieved when AUC = 1. To extended the global performance analysis we propose two ad hoc global metrics adapted from classical information retrieval metrics. We reglobal fer to these two metrics as

recall and *global f-measure*. Let the set A(P) be the union of the results returned by a population of queries $P = \{q_1, q_2, \dots, q_n\}$:

$$A(P) = \bigcup_{q_i \in P} A_{q_i},$$

where A_{q_i} is defined in the usual way as the set of results returned by q_i . Given a query population P?N^Q and a topic t, we compute global recall as a function Global – Recall : $\mathbb{N}^Q \times T \to [0, 1]$ defined as follows:

Global–Recall
$$(P, t) = \frac{|A(P) \cap D_t|}{|D_t|}.$$

To conclude the global performance analysis we compute the global *f*-measure as the harmonic mean of the average precision at rank 10 and the global recall of a population of queries. Formally, given a population of queries $P?N^Q$ and a topic $t \in T$, the global *f*-measure is a function Global – F – measure : $\mathbb{N}^Q \times T \rightarrow [0, 1]$ defined as follows:

$$Global - F - measure(P, t)$$

$$= 2 \cdot \frac{\overline{Precision@10}(P, t) \cdot Global - Recall(P, t)}{\overline{Precision@10}(P, t) + Global - Recall(P, t)}$$

where $\overline{\text{Precision@10}}(P, t)$ is the arithmetic mean of the *precision at rank 10* evaluated across all queries in the population *P*:

$$\overline{\operatorname{Precision}@10}(P,t) = \frac{\sum_{q \in P} \operatorname{Precision}@10(q,t)}{|P|}$$

The *Global-F-measure* function returns a performance rating for a whole population of queries, allowing to assign an aggregated global score to each of the evaluated strategies.

4.2. Performance analysis

In this section, we will evaluate the diversity strategies using different criteria. Tables 1, 2 and 3 summarize the figures used in each analysis. It is important to remark that figures whose numbering starts with "S" are provided in the supplementary material. These figures are not included in the discussion of performance analysis, but they show complementary results to those reported in this article.

Table 1 provides an overview of the figures related to performance evaluations reported for the 15 single-population strategies generated by the combination of the different fitness functions and parameter settings for two topics (BODY PAINTING and BIOINFORMATICS). The goal of these experiments is to contrast the different alternatives in terms of the *Precision@10* and *Recall* metrics. Table 1 also provides an overview of the figures showing the Pareto Fronts (PFs) achieved by the reported strategies.

Figs. 2 and 3 show the evolution of the *Precision@10* and *Recall* metrics during the training process. In both cases, the configurations

Table 1

An overview of the figures reporting the performance of the different strategies in terms of *Precision@10* and *Recall*, as well as the Pareto Fronts achieved by these strategies using evolution objectives Co1, Co2 and Co3.

| Metric | Training | Testing |
|---------------------|----------|---------|
| P@10 | Fig. 2 | Fig. 5 |
| Recall | Fig. 3 | Fig. 6 |
| Both (Pareto Front) | Fig. 4 | Fig. 7 |
| P@10 (Error Bars) | Fig. S1a | Fig. 8a |
| Recall (Error Bars) | Fig. S1b | Fig. 8b |

with hyper-mutation overcome the other parameterizations. Regarding the alternative fitness functions, the use of retrospective evaluations (Co2 and Co3) achieves the best performance in terms of *Precision*@10, but the classical approach (Co1) obtains the best results in terms of *Recall*. Nevertheless, the retrospective metrics achieve a quite reasonable evolution of their *Recall* values in the presence of high mutation rates. In contrast, Co1 has an erratic behavior for the *Precision*@10 values, which become decreasing after reaching the 50th generation. Therefore, from a general viewpoint, the retrospective metrics are preferable over the classical ones. Fig. 4 shows the PFs evaluated for the *Precision@10* and *Recall* objectives. In particular, the PFs of maximum AUC obtained by the different combinations of fitness functions and parameter settings are reported. In all cases, the best PFs are obtained with high levels of mutation (SuperM, HyperM, and HypoC+SuperM).

Figs. 5 and 6 show the evolution of the *Precision@10* and *Recall* metrics during the testing process, whereas Fig. 7 depicts the PFs of maximum AUC obtained during the testing stage. These results are analogous to those achieved from the training datasets, showing the benefits of using high levels of mutation. Fig. 8 illustrates with box plots the confidence intervals at the 95% level for the mean values of *Precision@10* and *Recall* for the last generation of evolved queries evaluated on the testing data. An important finding that derives from observing the performance of the queries on the testing set is that the MOEAs are not overfitting the training data.

Summarizing from these figures, it is clear that the retrospective fitness functions, as well as high levels of mutation contribute to a better evolution of the genetic algorithm in terms of the classical metrics used in information retrieval (*Precision@10* and *Recall*). An interesting observation from the analysis of Fig. 2 is the superiority in terms of *Precision@10* of Co2 and Co3 over Co1 in most of the parameterization settings. This is surprising, especially in the case of Co3, where *Preci*-



Fig. 2. The evolution of *Precision@10* for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations Co1, Co2, and Co3. Each point represents the mean values across the runs taking the average population performance.



Fig. 3. The evolution of *Recall* for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations Co1, Co2, and Co3. Each point represents the mean values across the runs taking the average population performance.

Pareto fronts for objectives *Precision@10* and *Recall* evaluated on the training set.



Fig. 4. Pareto fronts of a single population of queries evaluated on the training set for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations Co1 (left), Co2 (center) and Co3 (right). The reported results correspond to the runs with Pareto fronts of maximum AUC on the objective space *Precision@10* vs. *Recall*.

Precision@10 - Queries at different generations evaluated on the test set.



Fig. 5. Evaluating *Precision@10* on the test set using the queries evolved for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations Co1, Co2, and Co3. Each point represents the mean values across the runs taking the average population performance.

sion@10 is not an objective to be maximized. Nevertheless, we should take into account that *Precision* and *Recall* are normally competing objectives. Since Co1 has *Recall* as one of the two objectives to maximize and both Co2 and Co3 attempt to maximize *Retro-Recall* instead of *Recall*, we conjecture that an explanation for the observed behavior is that *Precision@10* (for the case of Co2) and *Retro-Precision@10* (for the case of Co3) undergo less degree of competition with *Retro-Recall* than with the classical *Recall* metric.

Another possible explanation for the superiority of the retrospective fitness functions over the classical ones is that the formers effectively achieve greater diversity (as will be analyzed in the next section). This result in a better exploration of the solution space, avoiding premature convergence and having a positive impact on the precision attained by the population of individuals.

4.3. Diversity preservation analysis

Table 2 contains an outline of the figures presented for contrasting the performance of the different single- and multi-population evolutionary strategies in terms of *Similarity*. It also contains an outline of the figures that illustrate the PFs for multiple populations of queries. The main goal of these experiments is evaluating the impact of several strategies on the diversity preservation of the populations.

Figs. 9 and 10 show the evolution of the *Similarity* metric for the different combinations of fitness functions and parameter settings using single and multiple populations respectively. An increase of *Similarity* represents a loss of genotypic diversity in the populations. Normally, every evolutionary algorithm presents a decreasing genotypic diversity when is converging to the optimal regions of the search space.



Fig. 6. Evaluating *Recall* on the test set using the queries evolved for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations Co1, Co2, and Co3. Each point represents the mean values across the runs taking the average population performance.

Pareto fronts for objectives Precision@10 and Recall evaluated on the test set.



Fig. 7. Pareto fronts of a single population of queries evaluated on the test set for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations Co1 (left), Co2 (center), and Co3 (right). The reported results correspond to the runs with Pareto fronts of maximum AUC on the objective space *Precision@10* vs. *Recall*.

However, this drop in the levels of diversity should occur slowly and gradually over the first generations of evolution. This desirable behavior is exhibited in Figs. 9d and 10d, where the fitness functions with retrospective metrics reach the lower values of *Similarity*. Regarding the comparison between population models, it is clear that the use of multiple populations significantly improves the genotypic diversity.

Figs. 11 and 12 show the PFs evaluated for the *Precision@10* and *Recall* goals, using multi-population strategies, from the training and testing datasets respectively. In particular, the PFs of maximum AUC obtained by the different combinations of fitness functions and parameter settings are reported. In all cases, the best PFs are obtained with high levels of mutation (HyperM). These results are consistent with the basic precepts of evolutionary computing, which state that high levels of mutation population promote the preservation of diversity.

4.4. Global performance analysis

Table 3 presents an overview of the figures related to global performance evaluations reported for the single- and multi-population strategies generated by the combinations of the different fitness functions and parameter settings. These plots illustrate the concept of population fitness by computing the global metrics, defined above, over the complete set of queries that integrate a single population and multiple populations.

Figs. 13 and 14 show the maximum, average and minimum values of AUC obtained for the different combinations of fitness functions and parameterizations from the training and testing datasets respectively. The reported results correspond to the PF of maximum AUC obtained in the different runs. In all cases, the parameterizations with high values of mutation achieve the best performance. With regard to



Fig. 8. Confidence intervals at the 95% level for the mean values of *Precision@10* and *Recall* evaluated on the test set using the queries evolved for the topics BODY PAINTING (left) and BIOINFORMATICS (right). Each point represents the mean values across the runs taking the average population performance for the last generation of queries and the bars show the corresponding confidence intervals.

Table 2

An overview of the figures contrasting the performance of the different strategies in terms of *Similarity* and the Pareto Fronts achieved by the multi-population strategies using evolution objectives Co1, Co2 and Co3.

| Metric | Set | Single pop | Multi-pop |
|-----------------------|----------|------------|-----------|
| Similarity | | Fig. 9 | Fig. 10 |
| Max AUC | Training | - | Fig. 11 |
| Max AUC | Testing | - | Fig. 12 |
| Similarity Error Bars | | Fig. S2 | Fig. S3 |
| Max Cardinality | Training | Fig. S4 | Fig. S5 |
| Max Cardinality | Testing | Fig. S6 | Fig. S7 |

the fitness functions, in the case of the training set, the alternative based on the use of classical metrics (Co1) achieves better coverage (AUC) that the functions based on retrospective metrics. However, when queries are evaluated on the test set, the three combinations (Co1, Co2, and Co3) show a similar behavior.

We should point out that while the PFs for single populations are obtained for populations of 100 individuals, the PFs for multiple populations are obtained by looking at super populations of size 1000 (10 sets of single populations, each containing 100 individuals). Typically, the size of the population affects the PF (with an increased AUC value for larger populations). In some way, this explains why the maximum, average and minimum AUC values across runs for the multi-population strategies are superior to the values achieved by the single-population strategies. This result is a good indicator of the advantage of using multiple populations rather than single populations. However, we should mention that this advantage not only results from having super populations with a larger number of individuals for the multi-population case but from the fact that the subpopulations that comprise each super population have evolved independently one from the other, attaining a higher coverage when jointly creating a Pareto Front.

Fig. 15 shows the means and confidence intervals at the 95% level for the last generation of evolved queries evaluated on the testing data for the *Global-Recall* metric. As it can be observed, higher mutation rates reach better results in terms of this metric and, in particular, the Co1 fitness functions combination seems to be better than Co2 and Co3. On the other hand, with respect to the single and multi-population strategies, the last ones show higher values regardless the crossover and mutation setup.

Figs. 16 and 17 show the evolution of *Global-F-measure* for the different combinations of fitness functions and parameter settings during the training phase, using single and multiple populations respectively. For both topics, the best performances are obtained using high mutation and retrospective recall. In particular, the classical metrics (Co1) have a poor evolution pattern, decreasing their values after the 50th generation.

Figs. 18 and 19 show the evolution of *Global-F-measure* for the different combinations of fitness functions and parameter settings during the testing stage, using single and multiple populations respectively. The conclusions about the evolution of *Global-F-measure* are similar to those obtained for the training datasets. In these plots, Co3 achieves the best performance using hyper-mutation.

Finally, Fig. 20 shows the means and confidence intervals at the 95% level for the *Global-F-measure* metric. As it was expected from the corresponding evolution figures, higher mutation rates in combina-



Fig. 9. Evaluating the *similarity* among queries at different generations using a **single population of queries** for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations **Co1**, **Co2**, and **Co3**. Each point represents the mean values across the runs taking the average population performance.





Fig. 10. Evaluating the *similarity* among queries at different generations using **multiple populations of queries** for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations **Co1**, **Co2**, and **Co3**. Each point represents the mean values across the runs taking the average population performance.

tion with *Retro-Recall* reach a much superior performance than the Co1 option.

Global-F-measure combines *Precision@10* and *Global-Recall* and therefore it provides a key score for each of the evaluated strategies as it integrates the assessment of the two main objectives of this research study. In other words, the *Global-F-measure* values achieved on the test set summarize the ultimate effectiveness of each of the evaluated strategies in simultaneously attaining a good coverage of relevant results at a global level and preserving accuracy at the local level. The fact that Co2 and Co3 are consistently superior to Co1 in terms of *Global-F-measure* (showing a statistically significant superiority in most cases) is a good indicator of the advantages of the retrospective fitness functions for preserving diversity without losing precision.

It is also interesting to note that Co2 and Co3 tend to achieve similar performance (not only in terms of *Global-F-measure* but also in terms of other performance analysis measures reported in this article). This may be because those queries that retrieve results that were previously retrieved by other queries are penalized in a similar way by the retrospective fitness functions used in Co2 and Co3. We conjecture that the effect of penalizing a query by *Retro-Recall* only (as is the case in Co2) might be almost identical to the effect of penalizing the same

query in terms both *Retro-Precision@10* and *Retro-Recall* (as is the case in Co3).

A final analysis of the *Global-F-measure* values on the test set (Fig. 20) suggests that the best strategy corresponds to the use of multiple populations, combined with low levels of crossover and high levels of mutation (HypoC+SuperM) and to the adoption of *Retro-Precision@10* and *Retro-Recall* as the objectives to be maximized (Co3). However, there is not a statistically significant superiority between this specific strategy and other reported strategies that take a multi-population approach with high levels of mutation (SuperM, HyperM, or HypoC+SuperM) as long as at least one of the retrospective fitness functions is used (Co2 or Co3).

5. Conclusion and future work

This work provides a systematic study of different computational strategies aimed at dealing with the diversity preservation problem in topic-based search, which constitutes a particular case of Big Data application. Topical search is formulated as a multi-objective optimization problem, where the optimal solutions are approximated by evolving populations of topical queries from an initial description of the topic of

Pareto fronts for objectives Precision@10 and Recall - Multiple populations of queries evaluated on the

training set.



Fig. 11. Pareto fronts of multiple populations of queries evaluated on the training set for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations Co1 (left), Co2 (center), and Co3 (right). The reported results correspond to the runs with Pareto fronts of maximum AUC on the objective space *Precision@10* and *Recall*.

Pareto fronts for objectives Precision@10 and Recall - Multiple populations of queries evaluated on the test

set.



Fig. 12. Pareto fronts of multiple populations of queries evaluated on the test set for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) when the pair of objectives to be maximized are given by combinations Co1 (left), Co2 (center), and Co3 (right). The reported results correspond to the runs with Pareto fronts of maximum AUC on the objective space *Precision@10* and *Recall*.

interest. The adopted approach for learning topical queries and the derived strategies for attaining diversity are fully automatic and can be applied to retrieve content related to any topic for which a description (topic characterization) and a set of relevant documents (training set) are available. The studied strategies include the use of specially developed diversity preservation fitness functions to which we refer to as *retrospective fitness functions*. The use of lower crossover rates and higher mutation rates, as well as the use of multiple populations of queries, have also been studied as a means to attain diversity. All these strategies were evaluated individually and in combination with each other.

The evaluations focused on analyzing the performance of the different strategies in terms of classical IR evaluation metrics (*Precision*@10 and *Recall*), as well as in terms of diversity preservation indicators

Table 3

An overview of the figures reporting the global performance of the different strategies using evolution objectives Co1, Co2 and Co3.

| Metric | Figure | | | |
|--|---|--|---|--|
| | Single pop | | Multi-pop | |
| | Training | Testing | Training | Testing |
| AUC Bars Global <i>Recall</i> – Error Bars Global <i>F-measure</i> Global <i>F-measure</i> – Error Bars Global <i>Recall</i> | Fig. 13 Fig. S12a Fig. 16 Fig. S13a Fig. S8 | Fig. 14 Fig. 15a Fig. 18 Fig. 20a Fig. S10 | Fig. 13 Fig. S12b Fig. 17 Fig. S13b Fig. S9 | Fig. 14 Fig. 15b Fig. 19 Fig. 20b Fig. S11 |

such as query similarity and the spread of the Pareto fronts. The final assessment was completed by looking at global performance measures such as the AUC of the Pareto fronts, the *Global-Recall* attained by the whole population of queries and a specially defined global measure called *Global-F-measure*, which simultaneously accounts for *Precision@10* and *Global-Recall*. Our final assessment suggests that the most effective strategies are those that use retrospective fitness functions, apply high mutation rates and evolve multiple populations of queries rather than a single one.

A critical element in the analysis of evolutionary algorithms applied to the exploration of very large collections of documents is their suitability for parallelization [85,86]. In this sense, it is important to mention that the strategies based on the use multiple populations are easily parallelizable and can be naturally executed on a cluster of machines through models that automatically support parallelism, such as MapReduce [87].

In the future, we plan to investigate additional strategies, such as adaptive mutation [88] as a means to attain diversity. Adaptive mutation can be naturally integrated into our approach by monitoring population diversity and adjusting the mutating rate accordingly. We also plan to investigate the use of similarity among queries as an objective to be minimized, as well as the use of alternative fitness functions that attempt to favor diversity by applying strategies different from those proposed in this work. Regarding the multi-population strategy, our next step is to explore the use of co-evolution using island models [89]. This kind of co-evolution models favors the use of adaptive parameter settings and parallel and distributed computing [90], which are common practices when evolutionary computing methods are applied in Big Data domains.

Acknowledgements

This work is supported by CONICET (PIP 11220120100487 and PIP 11220120100471), MinCyT (PICT 2014-0624), and Universidad Nacional del Sur (PGI-UNS 24/N039 and PGI-UNS 24/N042).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.asoc.2017.11.016.

Average, maximum and minimum AUC for objectives *Precision@10*

and Recall - Evaluated on the training set.







Fig. 13. Maximum, average and minimum AUC evaluated on the training set for the topics BODY PAINTING (left) and BIOINFORMATICS (right) when the pair of objectives to be maximized are given by combinations Co1 (top), Co2 (middle), and Co3 (bottom). Each bar shows the mean value across the runs of the corresponding AUC.

Average, maximum and minimum AUC for objectives

Precision@10 and Recall - Evaluated on the test set.







Fig. 14. Maximum, average and minimum AUC evaluated on the test set for the topics BODY PAINTING (left) and BIOINFORMATICS (right) when the pair of objectives to be maximized are given by combinations Co1 (top), Co2 (middle), and Co3 (bottom). Each bar shows the mean value across the runs of the corresponding AUC.

Mean and Confidence intervals - Global-Recall.





Fig. 15. Confidence intervals at the 95% level for the mean values of *Global-Recall* evaluated on the test set using the queries evolved using a single population of queries (top) and multiple populations of queries (bottom) for the topics BODY PAINTING (left) and BIOINFORMATICS (right). Each point represents the mean value across the runs for the last generation of queries and the bars show the corresponding confidence intervals.



Global-F-measure - Evolution on the training set for a single population of queries.

Fig. 16. The evolution of *Global-F-measure* for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) using a single population of queries when the pair of objectives to be maximized are given by combinations Co1, Co2, and Co3. Each point represents the mean value across the runs.



Fig. 17. The evolution of *Global-F-measure* for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) using **multiple populations of queries** when the pair of objectives to be maximized are given by combinations **Co1**, **Co2**, and **Co3**. Each point represents the mean value across the runs.



Global-F-measure - Queries at different generations evaluated on the test set for a single population of

Fig. 18. Evaluating *Global-F-measure* on the test set using the queries evolved for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) using a single population of queries when the pair of objectives to be maximized are given by combinations Co1, Co2, and Co3. Each point represents the mean value across the runs.

Global-F-measure - Queries at different generations evaluated on the test set for multiple populations of



Fig. 19. Evaluating *Global-F-measure* on the test set using the queries evolved for the topics BODY PAINTING (top) and BIOINFORMATICS (bottom) using **multiple populations of** queries when the pair of objectives to be maximized are given by combinations **Co1**, **Co2**, and **Co3**. Each point represents the mean value across the runs.

Mean and Confidence intervals - Global-F-measure.





Fig. 20. Confidence intervals at the 95% level for the mean values of *Global-F-measure* evaluated on the test set using the queries evolved using a single population of queries (top) and multiple populations of queries (bottom) for the topics BODY PAINTING (left) and BIOINFORMATICS (right).

References

- P. Sondhi, R. Chandrasekar, R. Rounthwaite, Using query context models to construct topical search engines, In: Proceedings of the Third Symposium on Information Interaction in Context, ACM, 2010, pp. 75–84, https://doi.org/10. 1145/1840784.1840797.
- [2] R. Chandrasekar, P. Sondhi, R. Rounthwaite, R. Chandrasekar, P. Sondhi, R. Rounthwaite, Topical search engines and query context models, uS Patent 9,177,045. (03.11.15).
- [3] D. Leake, A. Maguitman, T. Reichherzer, Experience-based support for human-centered knowledge modeling, Knowl. Based Syst. 68 (2014) 77–87, https: //doi.org/10.1016/j.knosys.2014.01.013.
- [4] R. Guha, V. Gupta, V. Raghunathan, R. Srikant, User modeling for a personal assistant, In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, ACM, 2015, pp. 275–284, https://doi.org/10.1145/ 2684822.2685309.
- [5] Q. Zheng, Z. Wu, X. Cheng, L. Jiang, J. Liu, Learning to crawl deep web, Inf. Syst. 38 (6) (2013) 801–819, https://doi.org/10.1016/j.is.2013.02.001.
- [6] P.N. Bennett, R.W. White, W. Chu, S.T. Dumais, P. Bailey, F. Borisyuk, X. Cui, Modeling the impact of short-and long-term behavior on search personalization, In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2012, pp. 185–194, https:// doi.org/10.1145/2348283.2348312.
- [7] Z.M. Gutt, R.L. Mains, A. Paushkina, Automatic search query suggestions with search result suggestions from user history, US Patent 8,370,329. (05.02.13).

- [8] J. Ramer, A. Soroca, D. Doughty, J. Ramer, A. Soroca, D. Doughty, User characteristic influenced search results, US Patent 9,223,878. (29.12.15).
- [9] C. Chesñevar, A. Maguitman, E. Estevez, R. Brena, Integrating argumentation technologies and context-based search for intelligent processing of citizens' opinion in social media, In: Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance, ICEGOV '12, ACM, New York, NY, USA, 2012, pp. 166–170, https://doi.org/10.1145/2463728.2463762.
- [10] D. Lin, S. Li, D. Cao, Making intelligent business decisions by mining the implicit relation from bloggers' posts, Soft Comput. 14 (12) (2009) 1317–1327, https://doi.org/10.1007/s00500-009-0499-9.
- [11] J. Budzik, K.J. Hammond, L. Birnbaum, Information access in context, Knowl. Based Syst. 14 (1–2) (2001) 37–53, https://doi.org/10.1016/ S0950-7051(00)00105-2
- [12] D. Leake, A. Maguitman, T. Reichherzer, Exploiting rich context: an incremental approach to context-based web search, in: A. Dey, B. Kokinov, D. Leake, R. Turner (Eds.), Modeling and Using Context, Vol. 3554 of Lecture Notes in Computer Science, Springer, Berlin Heidelberg, 2005, pp. 254–267, https://doi. org/10.1007/11508373 19.
- [13] R. Mehrotra, Topics, tasks & beyond: learning representations for personalization, In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, ACM, 2015, pp. 459–464, https://doi.org/10.1145/ 2684822.2697037.
- [14] Y. Zhang, M. Zhang, Y. Liu, C. Tat-Seng, Y. Zhang, S. Ma, Task-based recommendation on a web-scale, In: IEEE International Conference on Big Data (Big Data), IEEE, 2015, pp. 827–836, https://doi.org/10.1109/BigData.2015.7363829.

- [15] G. Sanchita, D. Anindita, Evolutionary algorithm based techniques to handle big data, In: Techniques and Environments for Big Data Analysis, Springer, 2016, pp. 113–158, https://doi.org/10.1007/978-3-319-27520-8 7.
- [16] S. Sagnika, B.S.P. Mishra, S. Dehuri, Parallel ga in big data analysis, In: Techniques and Environments for Big Data Analysis, Springer, 2016, pp. 101–111, https://doi.org/10.1007/978-3-319-27520-8 6.
- [17] R.L. Cecchini, C.M. Lorenzetti, A.G. Maguitman, N.B. Brignole, Using genetic algorithms to evolve a population of topical queries, Inf. Process. Manag. 44 (6) (2008) 1863–1878, https://doi.org/10.1016/j.ipm.2007.12.012.
- [18] R.L. Cecchini, C.M. Lorenzetti, A.G. Maguitman, N.B. Brignole, Multi-objective evolutionary algorithms for context-based search, J. Am. Soc. Inf. Sci. Technol. 61 (6) (2010) 1258–1274, https://doi.org/10.1002/asi.21320.
- [19] T.H. Cormen, C.E. Leiserson, R.L. Rivest, Introduction to Algorithms, 1st ed., MIT Press, Cambridge, MA, USA, 1990.
- [20] J.H. Holland, Adaptation in natural and artificial systems, Bradford Series in Complex Adaptive Systems, The University of Michigan Press, Ann Arbor, MI, USA, 1975.
- [21] A.E. Eiben, J.E. Smith, Introduction to Evolutionary Computing, Natural Computing Series, 1st ed., Springer, Berlin/Heidelberg, 2003.
- [22] R.L.T. Santos, C. Macdonald, I. Ounis, Search result diversification, Found. Trends Inf. Retr. 9 (1) (2015) 1–90, https://doi.org/10.1561/1500000040.
- [23] M.L. Mauldin, Maintaining diversity in genetic search, in: R.J. Brachman (Ed.), Proceedings of the National Conference on Artificial Intelligence, AAAI Press, 1984, pp. 247–250 http://www.aaai.org/Library/AAAI/1984/aaai84-050.php.
- [24] C. Lorenzetti, A. Maguitman, D. Leake, F. Menczer, T. Reichherzer, Mining for topics to suggest knowledge model extensions, ACM Trans. Knowl. Discov. Data 11 (2) (2016) 23:1–23:30, https://doi.org/10.1145/2997657
- [25] R. Qian, K. Zhang, G. Zhao, A topic-specific web crawler based on content and structure mining, In: 3rd International Conference on Computer Science and Network Technology (ICCSNT), IEEE, 2013, pp. 458–461.
- [26] D.A. Hanauer, Q. Mei, J. Law, R. Khanna, K. Zheng, Supporting information retrieval from electronic health records: a report of university of Michigan's nine-year experience in developing and using the electronic medical record search engine (emerse), J. Biomed. Inf. 55 (2015) 290–300.
- [27] A. Maguitman, D. Leake, T. Reichherzer, Suggesting novel but related topics: towards context-based support for knowledge model extension, In: IUI '05: Proceedings of the 10th International Conference on Intelligent User Interfaces, ACM Press, New York, NY, USA, 2005, pp. 207–214, https://doi.org/10.1145/ 1040830.1040876.
- [28] R. Kraft, C.C. Chang, F. Maghoul, R. Kumar, Searching with context, In: WWW '06: Proceedings of the 15th International Conference on World Wide Web, ACM, New York, NY, USA, 2006, pp. 477–486, https://doi.org/10.1145/ 1135777.1135847.
- [29] C.M. Lorenzetti, A.G. Maguitman, A semi-supervised incremental algorithm to automatically formulate topical queries, Inf. Sci. 179 (12) (2009) 1881–1892, https://doi.org/10.1016/j.ins.2009.01.029, (including Special Issue on Web Search.).
- [30] A. Fuxman, P. Pantel, Y. Lv, A. Chandra, P. Chilakamarri, M. Gamon, D. Hamilton, B. Kohlmeier, D. Narayanan, E. Papalexakis, B. Zhao, Contextual insights, In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, WWW Companion '14, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2014, pp. 265–266, https://doi.org/10.1145/2567948. 2577280.
- [31] B. Rhodes, T. Starner, The remembrance agent: a continuously running automated information retrieval system, Proceedings of The First International Conference on The Practical Application of Intelligent Agents and Multi Agent Technology (PAAM '96), AAAI Press, London, UK, 1996487–495.
- [32] A. Maguitman, D. Leake, T. Reichherzer, F. Menczer, Dynamic Extraction of Topic Descriptors and Discriminators: Towards automatic context-based topic search, In: Proceedings of the Thirteenth Conference on Information and Knowledge Management (CIKM), ACM Press, Washington, DC, 2004, pp. 463–472, https://doi.org/10.1145/1031171.1031260.
- [33] J.-J. Yang, R. Korfhage, Query optimization in information retrieval using genetic algorithms, In: Proceedings of the 5th International Conference on Genetic Algorithms, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993, pp. 603–613.
- [34] F.E. Petry, B.P. Buckles, D. Prabhu, D.H. Kraft, Fuzzy information retrieval using genetic algorithms and relevance feedback, in: S. Bonzi (Ed.), Proceedings of the 56th Annual Meeting of the American Society for Information Science (ASIS), vol. 30, Learned Information, Medford, NJ, USA, 1993, pp. 122–125.
- [35] H. Chen, G. Shankaranarayanan, L. She, A. Iyer, A machine learning approach to inductive query by examples: an experiment using relevance feedback, ID3, genetic algorithms, and simulated annealing, J. Am. Soc. Inf.

Sci. 49 (8) (1998) 693–705, https://doi.org/10.1002/ (SICI)1097-4571(199806)49:8<693::AID-ASI4>3.0.CO;2-O.

- [36] O. Cordón, F. Moya, C. Zarco, Automatic learning of multiple extended Boolean queries by multiobjective gap algorithms, in: V. Loia, M. Nikravesh, L.A. Zadeh (Eds.), Fuzzy Logic and the Internet, Vol. 137 of Studies in Fuzziness and Soft Computing, Springer, 2004, p. 47, https://doi.org/10.1007/978-3-540-39988-9_3.
- [37] A.G. López-Herrera, E. Herrera-Viedma, F. Herrera, A study of the use of multi-objective evolutionary algorithms to learn Boolean queries: a comparative study, J. Am. Soc. Inf. Sci. Technol. 60 (6) (2009) 1192–1207, https://doi.org/10. 1002/asi.v60:6.
- [38] M. Caramia, G. Felici, A. Pezzoli, Improving search results with data mining in a thematic search engine, Comput. Oper. Res. 31 (14) (2004) 2387–2404, https:// doi.org/10.1016/S0305-0548(03)00194-1.
- [39] P. Malo, P. Siitari, A. Sinha, Automated query learning with Wikipedia and genetic programming, Artif. Intell. 194 (2013) 86–110, https://doi.org/10.1016/j. artint.2012.06.006.
- [40] P. Bhatnagar, N. Pareek, Improving pseudo relevance feedback based query expansion using genetic fuzzy approach and semantic similarity notion, J. Inf. Sci. 40 (4) (2014) 523–537, https://doi.org/10.1177/0165551514533771.
- [41] C. Carpineto, M. D'Amico, G. Romano, Evaluating subtopic retrieval methods: clustering versus diversification of search results, Inf. Process. Manag. 48 (2) (2012) 358–373, https://doi.org/10.1016/j.ipm.2011.08.004.
- [42] J. Carbonell, J. Goldstein, The use of mmr, diversity-based reranking for reordering documents and producing summaries, In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR' 98, ACM, New York, NY, USA, 1998, pp. 335–336, https://doi.org/10.1145/290941.291025.
- [43] R. Agrawal, S. Gollapudi, A. Halverson, S. Ieong, Diversifying search results, In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, WSDM '09, ACM, New York, NY, USA, 2009, pp. 5–14, https://doi.org/10.1145/1498759.1498766.
- [44] R.L. Santos, C. Macdonald, I. Ounis, Exploiting query reformulations for web search result diversification, In: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, 2010, pp. 881–890, https://doi.org/10.1145/1772690.1772780.
- [45] A. Bouchoucha, J. He, J.-Y. Nie, Diversified query expansion using concept net, In: Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management, CIKM '13, ACM, New York, NY, USA, 2013, pp. 1861–1864, https://doi.org/10.1145/2505515.2507881.
- [46] H. Ma, M.R. Lyu, I. King, Diversifying query suggestion results, in: M. Fox, D. Poole (Eds.), Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI Press, 2010, pp. 1399–1404 http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1827.
- [47] H. Hu, M. Zhang, Z. He, P. Wang, W. Wang, Diversifying query suggestions by using topics from Wikipedia, In: Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) – Vol. 01, WI-IAT '13, IEEE Computer Society, Washington, DC, USA, 2013, pp. 139–146, https://doi.org/10.1109/WI-IA.T.2013.21.
- [48] H. Zheng, J. Zhao, Y. Zhang, Y. Jiang, S. Xia, An ontology-based approach to query suggestion diversification, in: C.K. Loo, K.S. Yap, K.W. Wong, A.T.B. Jin, K. Huang (Eds.), Neural Information Processing – 21st International Conference, ICONIP 2014, Proceedings, Part II, Vol. 8835 of Lecture Notes in Computer Science, Springer, 2014, pp. 437–444, https://doi.org/10.1007/ 978-3-319-12640-1_53.
- [49] Y. Kim, W.B. Croft, Diversifying query suggestions based on query documents, In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14, ACM, New York, NY, USA, 2014, pp. 891–894, https://doi.org/10.1145/2600428.2609467.
- [50] C. Darwin, On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, John Murray, London, UK, 1859.
- [51] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.
- [52] M. Srinivas, L.M. Patnaik, Genetic algorithms: a survey, Computer 27 (6) (1994) 17–26, https://doi.org/10.1109/2.294849.
- [53] M. Mitchell, An Introduction to Genetic Algorithms, MIT Press, Cambridge, MA, USA, 1996.
- [54] G.T. Pulido, Optimización Multiobjetivo Usando un Micro Algoritmo Genético, Maestría en Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, México, 2001, September, (Master's thesis).

- [55] F. Lin, G. He, An improved genetic algorithm for multi-objective optimization, In: Proceedings of the Sixth International Conference on Parallel and Distributed Computing Applications and Technologies (PDCAT), IEEE Computer Society, Los Alamitos, CA, USA, 2005, pp. 938–940, https://doi.org/10.1109/PDCAT. 2005.84.
- [56] C.A. Coello Coello, G.B. Lamont, D.A. Van Veldhuiz, Evolutionary algorithms for solving multi-objective problems, Genetic and Evolutionary Computation, 2nd ed., Springer-Verlag New York, Inc, New York, NY, 2007https://doi.org/10. 1007/978-0-387-36797-2.
- [57] K. Deb, Multi-Objective Optimization Using Evolutionary Algorithms, John Wiley & Sons, Ltd., Chichester, W Sussex, UK, 2001.
- [58] V. Pareto, Cours d'Economie Politique, Droz, Genève, 1896.
- [59] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evolut. Comput. 6 (2) (2002) 182–197.
- [60] S. Bleuler, M. Laumanns, L. Thiele, E. Zitzler, PISA a platform and programming language independent interface for search algorithms, in: C.M. Fonseca, P.J. Fleming, E. Zitzler, K. Deb, L. Thiele (Eds.), Evolutionary Multi-Criterion Optimization, Vol. 2632 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2003, pp. 494–508, https://doi.org/10.1007/3-540-36970-8_35.
- [61] A. Toffolo, E. Benini, Genetic diversity as an objective in multi-objective evolutionary algorithms, Evolut. Comput. 11 (2) (2003) 151–167, https://doi.org/10. 1162/106365603766646816.
- [62] K. Tahernezhadiani, A. Hamzeh, S. Hashemi, Towards enhancing solution space diversity in multi-objective optimization: a hypervolume-based approach, Int. J. Artif. Intell. Appl. 3 (1) (2012) 65–81, https://doi.org/10.5121/ijaia.2012.3106.
- [63] J.B. Ghosh, Computational aspects of the maximum diversity problem, Oper. Res. Lett. 19 (4) (1996) 175–181.
- [64] M.G. Resende, R. Martí, M. Gallego, A. Duarte, Grasp and path relinking for the max-min diversity problem, Comput. Oper. Res. 37 (3) (2010) 498–508.
- [65] R.L. Hagin, Investment management: portfolio diversification, risk, and timing-fact and fiction, Vol. 235, John Wiley & Sons, 2004.
- [66] O.A. Prokopyev, N. Kong, D.L. Martinez-Torres, The equitable dispersion problem, Eur. J. Oper. Res. 197 (1) (2009) 59–67.
- [67] K.D.V. Escobar, F. López-Pires, B. Barán, F. Sandoya, Multi-objective maximum diversity problem, In: XLIII Latin American Computing Conference, CLEI 2017, September 4–8, Córdoba, Argentina, 2017.
- [68] K.A. De Jong, An analysis of the behavior of a class of genetic adaptive systems, University of Michigan, Ann Arbor, MI, USA, 1975, (Ph.D. thesis).
- [69] D.E. Goldberg, J. Richardson, Genetic algorithms with sharing for multimodal function optimization, Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application, L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 198741–49.
- [70] F. Menczer, M. Degeratu, W.N. Street, Efficient and scalable Pareto optimization by evolutionary local selection algorithms, Evolut. Comput. 8 (2) (2000) 223–247, https://doi.org/10.1162/106365600568185.
- [71] J.A. Markert, D.M. Champlin, R. Gutjahr-Gobell, J.S. Grear, A. Kuhn, T.J. Me-Greevy, A. Roth, M.J. Bagley, D.E. Nacci, Population genetic diversity and fitness in multiple environments, BMC Evolut. Biol. 10 (1) (2010) 205.
- [72] C.J. Goodnight, Experimental studies of community evolution II: the ecological basis of the response to community selection, Evolution 44 (6) (1990) 1625–1636, https://doi.org/10.2307/2409342.
- [73] C.J. Goodnight, Experimental studies of community evolution I: the response to selection at the community level, Evolution 44 (6) (1990) 1614–1624, https:// doi.org/10.2307/2409341.

- [74] P. Collet, E. Lutton, F. Raynal, M. Schoenauer, Polar IFS + parisian genetic programming = efficient IFS inverse problem solving, Genet. Program. Evol. Mach. 1 (4) (2000) 339–361, https://doi.org/10.1023/A:1010065123132.
- [75] D. Gupta, S. Ghafir, An overview of methods maintaining diversity in genetic algorithms, Int. J. Emerg. Technol. Adv. Eng. 2 (5) (2012) 56–60.
- [76] J. Shi, M. Gong, W. Ma, L. Jiao, A multipopulation coevolutionary strategy for multiobjective immune algorithm, Sci. World J. (2014) https://doi.org/10.1155/ 2014/539128.
- [77] J.K. Cochran, S.-M. Horng, J.W. Fowler, A multi-population genetic algorithm to solve multi-objective scheduling problems for parallel machines, Comput. Oper. Res. 30 (7) (2003) 1087–1102, https://doi.org/10.1016/ S0305-0548(02)00059-X.
- [78] G. Amati, Probabilistics Models for Information Retrieval based on Divergence from Randomness, Department of Computing Science, University of Glasgow, UK, June, 2003https://doi.org/10.1145/582415.582416, (Ph.D. thesis).
- [79] J.J. Rocchio, Relevance feedback in information retrieval, in: G. Salton (Ed.), The Smart Retrieval System – Experiments in Automatic Document Processing, Prentice-Hall, Englewood Cliffs, NJ, USA, 1971, pp. 313–323, (Ch. 14).
- [80] G. Salton, C. Yang, On the specification of term values in automatic indexing, J. Document. 29 (1973) 351–372, https://doi.org/10.1108/eb026562.
- [81] R.A. Baeza-Yates, B.A. Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley Longman Publishing Co., Inc., Reading, MA, USA, 1999.
- [82] D. Hawking, N. Craswell, Overview of the TREC-2001 Web track, in: E.M. Voorhees, D.K. Harman (Eds.), Proceedings of the Tenth Text Retrieval Conference TREC-2001, Vol. Special Publication 500-250, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA, 2001, pp. 61–67.
- [83] R.L. Cecchini, C.M. Lorenzetti, A.G. Maguitman, F. Menczer, A semantic framework for evaluating topical search methods, CLEI Electron. J. 14 (1) (2011) http://www.clei.cl/cleiej/paper.php?id=211.
- [84] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, C. Lioma, Terrier: a high performance and scalable information retrieval platform, In: Proceedings of ACM SIGIR'06 Workshop on Open Source Information Retrieval (OSIR 2006), ACM, New York, NY, USA, 2006.
- [85] A. Umbarkar, M. Joshi, Review of parallel genetic algorithm based on computing paradigm and diversity in search space, ICTACT J. Soft Comput. 3 (2013) 615–622.
- [86] Y.-J. Gong, W.-N. Chen, Z.-H. Zhan, J. Zhang, Y. Li, Q. Zhang, J.-J. Li, Distributed evolutionary algorithms and their models: a survey of the state-of-the-art, Appl. Soft Comput. 34 (2015) 286–300, https://doi.org/10.1016/j.asoc.2015.04. 061.
- [87] J. Dean, S. Ghemawat, Mapreduce: a flexible data processing tool, Commun. ACM 53 (1) (2010) 72–77, https://doi.org/10.1145/1629175.1629198.
- [88] M. Srinivas, L.M. Patnaik, Adaptive probabilities of crossover and mutation in genetic algorithms, IEEE Trans. Syst. Man Cybern. 24 (4) (1994) 656–667, https://doi.org/10.1109/21.286385.
- [89] R. Tanese, Distributed genetic algorithms, in: J.D. Schaffer (Ed.), Proceedings of the Third International Conference on Genetic Algorithms, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1989, pp. 434–439.
- [90] F. Asteasuain, J.A. Carballido, G.E. Vazquez, I. Ponzoni, in: J.S. Sichman, H. Coelho, S. Oliveira Rezende (Eds.), Advances in Artificial Intelligence IB-ERAMIA-SBIA 2006, Vol. 4140 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Berlin, 2006, pp. 188–197, https://doi.org/10.1007/11874850_23.