

Fuzzy Logic Ranking for Personalized Geographic Information Retrieval

Garnett Wilson, Rodolphe Devillers, and Orland Hoerber

Abstract This work describes a novel fuzzy logic system designed to meet the real world demand of providing intelligent ranking to large repositories of documents previously encoded with non-fuzzy (crisp) metadata. The fuzzy logic prototype was tested in practice to complement the GeoConnections Discovery Portal, which is a web portal for specialized search and retrieval of Canadian geographic data resources via an associated web service. Users of the portal are able to query the system and then filter their search results by selecting topic categories, spatial and temporal extents, and resource types. The authors present a fuzzy logic information retrieval system that utilizes document metadata, and compare it to an unranked listing, standard term frequency-inverse document frequency (TF-IDF) ranking, and a TF-IDF/fuzzy hybrid system. Results indicate that the fuzzy logic system provided the overall highest precision among the top ranked documents for searches by an expert user, and that these results were robust with respect to the number of results returned by a number of different query types.

1 Introduction

Many fuzzy logic systems that aim to improve the ranking of the search results require existing specialized fuzzy metadata [3, 8] or the creation of supporting fuzzy

Garnett Wilson
Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada
e-mail: gwilson@mun.ca

Rodolphe Devillers
Department of Geography, Memorial University of Newfoundland, St. John's, NL, Canada
e-mail: rdeville@mun.ca

Orland Hoerber
Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada
e-mail: hoerber@mun.ca

ontologies [5, 7]. However, large repositories, such as the ones accessed in this work, are often encoded with crisp metadata. The fuzzy logic system we present in this work is designed to take advantage of XML encodings in large online repositories of existing documents that do not contain fuzzy data. Design choices for the system were guided by an actual end user need to intelligently rank these documents using existing metadata, resulting in the design of a novel fuzzy logic solution that uses existing non-fuzzy (crisp) ontologies.

The GeoConnections Discovery Portal ¹ provides search capabilities for geospatial data resources contained within multiple repositories. However, the search facilities simply provide a filtering mechanism, without any ranking capabilities. The web portal has a connected web service that can be used to return XML documents that contain the supplied search terms. As a research project in collaboration with GeoConnections, we created a system that implements an intelligent ranking of the documents returned by the web service. To this end, we chose to provide a personalized ranking based on per session user profiles because a typical end user of the portal may search for information on many unrelated topics. The performance of the fuzzy logic retrieval system is compared to unranked documents returned by the web service, the standard information retrieval ranking of term frequency-inverse document frequency (TF-IDF) [1], and a hybrid TF-IDF/Fuzzy system.

2 Related Work

The traditional approach to ranking search results has been to order documents that match the search query based on measures such as term frequency-inverse document frequency (TF-IDF) [1]. Such an approach gives ranking preference to query terms that appear frequently within the search results, but are rare with respect to their use across all of the documents in the collection. Although such ranking methods can be quite effective, there are opportunities for improvement. One such improvement that we explore within this research is to personalize the ranking of the search results based on additional information and preferences regarding the searcher's needs [9].

Castellano et al. [3] (and related work in Mencar et al. [8]) have described modelling of user preferences with fuzzy profiles. In particular, their work focuses on a system where individual users have multiple (often diverging) sets of interests in a user profile. An adaptive algorithm is designed to learn these multiple profiles of users and thus personalize their search results. For the practical purposes of a user researching a particular area of interest in one setting (for an individual professional contract or student project, for example) it made little sense in our implementation to keep likely unrelated user profiles from past searches. For instance, if a user is doing a search regarding road infrastructure, the search on caribou habitat that they performed last week would inhibit effective ranking. Our work uses fuzzy logic foundations behind the modelling of user profiles (as in [3]). However, the method

¹ <http://geodiscover.cgdi.ca>

of ranking in the fuzzy system used in this work is significantly different and makes use of existing non-fuzzy metadata in documents, rather than requiring the *a priori* encoding of fuzzy data within all metadata as in [3, 8].

There is a large body of work dedicated to geographic information retrieval (GIR), where many works focus on metadata encoding, methods to determine whether or not an arbitrary query is of a geographic nature, or ways to combine geographical information (such as location) with other geographical information (such as a city being known to be in a country) [6]. The scope of this work is different, and looks at practical decisions for a fuzzy logic ranking system that is to provide improved ranking based on metadata that happens to be geographic in nature (but can be applied to other domains). A GIR system by Leite [7] that does use geographic fuzzy relations differs significantly from our study because it uses a pre-existing ontology for Brazilian territories and climate in order to expand user queries for better search. Our work does not expand user queries for improved search, and the substantial amount of existing metadata from the geographic web services has no fuzzy metadata encoding. Furthermore, our method is a practical solution that can be easily implemented over top of existing metadata repositories and does not require additional metadata or the development of a formal ontology.

In order to explore the methods for effective ranking of documents devised in this work, we examine a standard information retrieval ranking mechanism that ranked documents based on terms in the document (TF-IDF), a fuzzy logic system that relies solely on the metadata tags in the documents stored on the online databases, and a combination of the TF-IDF and fuzzy rankings. Other researchers have attempted to combine TF-IDF and fuzzy rankings. For instance, Holi et al. [5] combine a fuzzy ontology with the TF-IDF measure itself. Rubens [12] combine traditional logic ranking with TF-IDF using fuzzy logic. Rather than using fuzzy logic within the TF-IDF framework, as in [5, 12], we combine the separate results of TF-IDF and fuzzy logic by placing an appropriate weight on each method. The aim of combining the weighted measures is to allow consideration of both the textual contents and metadata for effective ranking (discussed further in Section 4.3).

3 Fuzzy Logic Search Portal

In order to improve the relevance of the documents provided by the GeoConnections Discovery Portal, we further filter the documents retrieved from the portal's corresponding web service. We accomplish this by creating a web interface front end for our prototype that incorporates its own web server and other back end elements. The overall system architecture is shown in Figure 1. In particular, the public-facing web front end uses AJAX, JQuery, and sends JSON objects to a Jetty web server that allows Java classes to be used server-side for the processing of our document ranking techniques and then returned to the web portal. The Jetty web server communicates with the web service to retrieve unranked documents encoded with XML metadata from the online geography resources.

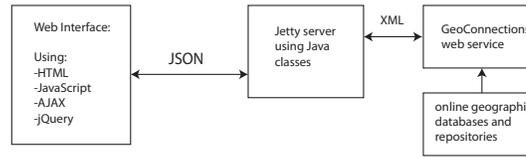


Fig. 1 Fuzzy logic retrieval prototype architecture

The web interface front end is shown in Figure 2. There is a query box in which the searcher can enter their textual query, and four buttons, each of which activates one of the four methods for ranking the search results (Unranked, TF-IDF, Fuzzy Logic, and TF-IDF/Fuzzy Logic). Titles of separate returned documents are displayed at the bottom of the page, and upon clicking the title the user can examine the corresponding document’s abstract. The interface has been developed for experimental purposes, with potential future interfaces based on the findings in this research. The User Profiles section of the web interface allow users to specify, on a scale of 1 to 10, how interested they are in each of 68 topics specified in the metadata of the documents. The metadata are not custom encoded for our fuzzy logic system, but we make creative use of the large amount of existing metadata in documents of a number of geographic databases provided by the web service. In particular, there are a number of topic categories under “Categories,” types of data resources under “Resource Type,” and political and physiographic locations under “Locations.” Examples of values for these metadata attributes used in our experiments will be provided in Section 5.

4 Ranking Methods

We establish a performance baseline using a standard TF-IDF ranking of documents. The standard TF-IDF is compared to a fuzzy logic implementation that takes the user specifications of interest from each of several topic categories, regions, and information resource types and returns rankings based on those user interests as encoded in the metadata. In addition, a hybrid of TF-IDF and fuzzy logic is examined. In this section, we specify the details of each ranking method implementation.

4.1 Term Frequency - Inverse Document Frequency (TF-IDF)

In information retrieval ranking based on term frequencies, each document d is considered as a vector of elements \vec{d} with each dimension corresponding to a term contained in that document d . A set of search query terms is considered a special

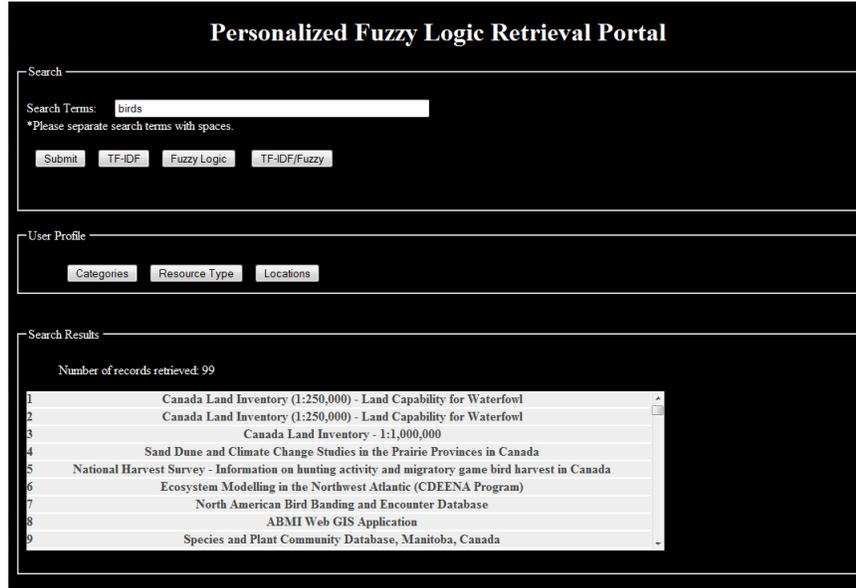


Fig. 2 The main page of the fuzzy logic retrieval portal

type of document vector, \vec{q} . In order to reduce the dimensionality of these vectors, stop word removal is used to eliminate common terms (prepositions, conjunctions, etc.) and Porter's stemming algorithm [10] strips prefixes and suffixes from terms, thereby counting terms with the same root together.

When representing the document (or query) as a vector, each dimension of the vector that corresponds to a term is assigned a weight, typically a TF-IDF weight. TF is a function of term frequency, and IDF is a function of inverse document frequency. The TF rewards a term occurring more frequently in a document, while IDF provides a lower weight to a term that appears in many documents. Let N be the total number of documents in the filtered collection based on the user-supplied query ², N_t be the number of those documents containing a term t , and $f_{t,d}$ be the number of occurrences of t in a document d . In this work the TF and IDF function are standards in information retrieval [1], where TF is defined as

$$TF = \begin{cases} \log(f_{t,d}) + 1 & \text{if } f_{t,d} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and IDF is defined as

$$IDF = \log(N/N_t). \quad (2)$$

TF-IDF is then the product of the term frequency and inverse document frequency,

² It is not possible to know the number, or content, of all documents in all repositories served by the web portal.

$$TF \cdot IDF. \quad (3)$$

To measure the similarity of each document vector \vec{d} to the query vector \vec{q} , the cosine similarity measure is used. The cosine similarity measure is actually the dot product of the document and query vectors normalized to unit length:

$$sim(\vec{d}, \vec{q}) = \frac{\vec{d}}{|\vec{d}|} \cdot \frac{\vec{q}}{|\vec{q}|}. \quad (4)$$

Since the components of the vectors are not negative, the cosine similarity measure ranges from 0 to 1. When the cosine similarity value is high, then the vectors of the document and query are similar to one another. Therefore, this measure of similarity can be used to rank the search results such that those that are most similar to the query are placed at the top of the search results list.

4.2 Fuzzy Logic

The fuzzy logic system considers user-specified values on a scale of 0 to 10 for each attribute a from the set of all possible attributes, A . The first step in the fuzzy logic implementation is to gather all the user-defined values for each attribute (ranked from 1 to 10, adjusted to a float in the interval $[0, 1]$) from the web site and send them to the Jetty web server for processing. More specifically, the collection of 68 metadata tags (total of possible values of all attributes) in a user profile is divided into three different attribute possibilities as indicated under “User Profile” on the web portal. These attributes are “Categories,” “Resource Type,” and “Locations,” so $|A| = 3$ and we have $A = a_1, a_2, a_3$. Furthermore, given the possible metadata tags in the documents, the 68 possible attributes are $|Dom(a_1)| = 18$, $|Dom(a_2)| = 17$, and $|Dom(a_3)| = 33$ where $Dom(a_i)$ is the set of all possible values of attribute a_i . The unranked documents are then retrieved from the web service using the Java implementation on the Jetty web server based on the search terms. The attributes corresponding to the metadata for the returned items, which is consistent across the collection of repositories, is then analyzed by the fuzzy logic system. For the fuzzy logic system, these documents will be referred to as “items” to distinguish from the TF-IDF ranking.

We adapt the notion of component-based fuzzy user profiles similar to those applied by Castellano in [3]. Other than this component of the fuzzy logic system, we create our own methodology for forming fuzzy metadata encoding from existing crisp metadata. In addition, we use a fuzzy similarity measure more appropriate for our information retrieval-focused purposes. Each item in a set of fuzzy metadata is expressed as a couple $\langle a, \mu \rangle$ where $a \in A$ and $\mu : Dom(a) \rightarrow [0, 1]$ is a fuzzy set defined on $Dom(a)$. An item that a user is interested in can then be described as

$$I = \{ \langle a, \mu \rangle \mid a \in A \} \quad (5)$$

with the constraint that each attribute does not occur more than once in a description.

User profiles represent the interests that each individual user has for the attributes of the returned items. User profiles are expressed as

$$P = \{ \langle a, \mu \rangle \mid a \in A \}. \quad (6)$$

Regarding items (documents) to be compatible with user profiles, each item I is not encoded as fuzzy metadata. To convert the existing metadata labels in documents, we use a fuzzy control to create appropriate fuzzy attribute value content. In particular, the total of all values in an item I corresponding to each of the three profile attributes a_1, a_2, a_3 is totaled, and each attribute value x is divided by the total of all x values for that attribute. More formally,

$$sum_{a_i}(I) = \sum_{x \in Dom(a_i)} \mu_I(x) \quad (7)$$

if μ_I is a fuzzy set of items. Each $x \in a_i$ is then replaced with x' where

$$x' = x / sum_{a_i}(I). \quad (8)$$

This production of fuzzy data from the existing metadata labeling means that if the article has more metadata tags labeling it under one of the three attributes of a user profile, each metadata tag has less value. This makes intuitive sense: If an article is labeled exclusively as one value under one of the user profile's attributes, and the user places importance on that value, then that item ought to be of higher relevance to the user than another article where it is labeled with several values under that same attribute. As a practical example of this, suppose one item in a search for "birds" is labeled as only "Biology and Ecology" under the Category attribute, and another item is labeled with all of "Atmosphere and Climate," "Biology and Ecology," and "Business and Economic." There will then be more emphasis on the former item if the user is very interested in "Biology and Ecology" and has little or no interest in atmospheric conditions effecting birds or associated economic issues.

The user profile P is a fuzzy set μ_P and the set of items returned from the web service prior to ranking is a fuzzy set of items, μ_I . In the fuzzy logic system, we instantiate all possible attribute values in each item I and user profile P , so each vector is of length 68 (all possible attribute values, the $Dom(A)$). If a user does not classify an attribute value as being of interest, it has a value of 0. Similarly, if the attribute value is not present in an item, it is given a value of 0 for that attribute. To calculate the compatibility degree between the fuzzy set of items μ_I and the user profile μ_P of attributes, we use the dot product of μ_P and μ_I :

$$K(\mu_I, \mu_P) = \mu_I \cdot \mu_P \quad (9)$$

The dot product is one of the standard functions used in engineering-based fuzzy systems, ranging from 0 (complete dissimilarity) to 1 (identical), and it is suitable for pairwise comparison of vector-based data points [11] (individual items and the

user profile). This dot product operation is essentially the same as used in the TF-IDF method in Equation 4 (only the vectors are based on entirely different information). Additionally, the vector returned by the dot product operation of the matrix of multiple items and the user profile vector is a list of the compatibilities (in the range 0 to 1) of each item to the user profile. These compatibilities are then used to rank the items as displayed to the end user on the web portal.

4.3 TF-IDF/Fuzzy Logic Hybrid

TF-IDF, considered in isolation, will provide ranking of documents based on the terms contained within them in relation to the user’s search terms. Fuzzy logic, considered in isolation, will provide ranking of document (items) based on the user’s ranking of particular topics contained in their metadata. Thus, to combine the information contained in both the text and metadata of a document (item in fuzzy logic), we use a hybrid ranking similarity measure

$$sim'(\vec{d}, \vec{q}, \mu_I, \mu_P) = \alpha \cdot sim(\vec{d}, \vec{q}) + \beta \cdot K(\mu_I, \mu_P) \quad (10)$$

where $sim(\vec{d}, \vec{q})$ and $K(\mu_I, \mu_P)$ are as defined in Equations 4 and 9, respectively. The coefficients α and β are used to put appropriate weights on the similarity and compatibility measures, respectively, are in the interval [0, 1] and add up to 1.0.

It is convenient to combine the similarity measure of the TF-IDF ranking and the compatibility measure of fuzzy logic into one measure of similarity since both measures are in the interval [0, 1] and there is a measure for each document (item). Due to the significant difference in the dimensionality of the document vectors used in calculating the TF-IDF ranking compared to the fuzzy vectors used in calculating the fuzzy ranking, we observe that typically the TF-IDF similarity measures are 2-3 orders of magnitude smaller than the compatibility measures between the fuzzy vectors and each item. In this work we therefore use weightings of $\alpha = 0.99$ and $\beta = 0.01$, which we found to provide a suitable balance between the two document (item) sets returned by TF-IDF and fuzzy measures.

5 Results

We assess the ranking of the documents returned from the web service (which is known to be unranked, and is designated “Unranked”), TF-IDF that is our baseline measure from standard information retrieval (“TF-IDF”), fuzzy logic ranking (“Fuzzy”), and the combination of TF-IDF and fuzzy logic as in Equation 10 (“TF-IDF/Fuzzy”). An expert user formulated six searches designed to test the effectiveness of the ranking mechanisms across a spectrum of specificity with respect to both spatial and thematic components of the search query. The searches chosen by the ex-

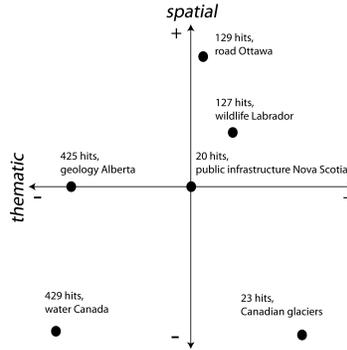


Fig. 3 Specificity of spatial and thematic component of expert searches

pert are plotted along the thematic and spatial spectra in Figure 3. The number of documents returned by the web service for each search query is also provided.

For each of these separate searches, we calculate the effectiveness of the rankings using precision. Recall is commonly used as an effectiveness measure with precision, giving the proportion of relevant documents in the result set. We do not use recall here because it is not possible to know the number of relevant documents in the large online databases served by the GeoConnections-associated web services. Precision measures the proportion of the result set that is relevant and is defined as

$$precision@k = \frac{|\text{Rel} \cap \text{Res}[1..k]|}{|\text{Res}[1..k]|} \tag{11}$$

for the top k documents returned by a search query [2]. The expert user judged relevance of a document based on whether or not the document contained useful information given his search based on both title (and abstract if necessary). The precision results for $k = 5, 10, 20$ are shown in Figure 4 below. The fuzzy logic user profile attribute values chosen by the expert are provided in Table 1.

The precision results in Figure 4 show that, as expected, all ranking methods outperformed the unranked search results. The fuzzy logic system performed as well as (or better than) the traditional TF-IDF ranking in all searches for the top $k = 20$ documents, and all but one instance (“Canadian glaciers”) for $k = 10$. This result clearly indicates the value of geographic metadata over document text for effective ranking of the results from the online repositories. Results at $k = 5$ provided no discernible trend when comparing TF-IDF and fuzzy solutions. As k increases, the precision decreases for almost all fuzzy searches, indicating that it provides an effective ranking mechanism with relevant documents high in the list and non-relevant documents lower in the list. Considering results at $k = 10$ and $k = 20$, as the spatial specificity increased from Canada (Searches e and f) to province (“Alberta” and “Nova Scotia”) or region in a province (“Labrador”), the fuzzy logic implementation provided the most benefit. However, for the most spatially specific search (Search c, involv-

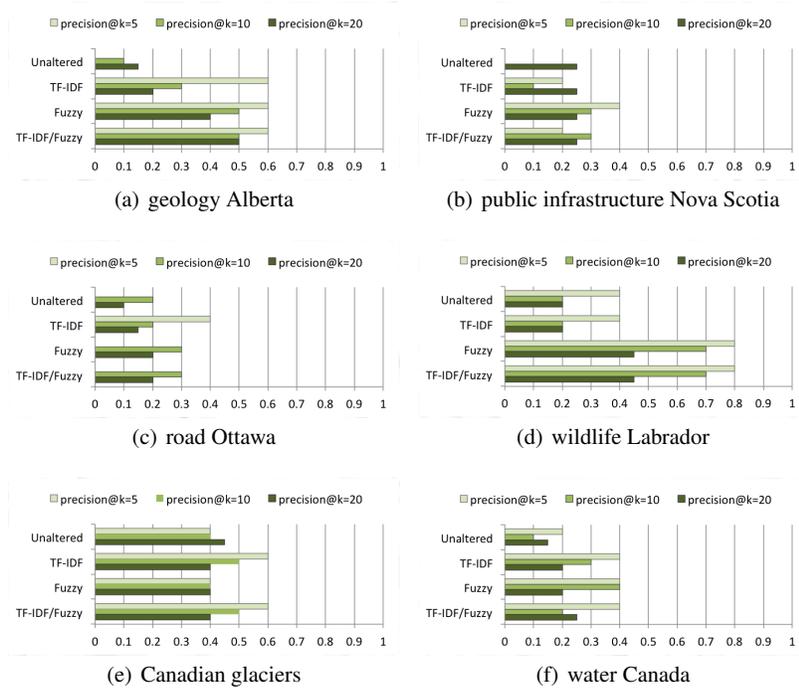


Fig. 4 Precision for top k documents returned per search query

ing the city “Ottawa”) the benefit of the fuzzy search was moderate for $k = 10$ and $k = 20$. In Search d, where the fuzzy logic search outperformed TF-IDF considerably, the expert user provided a detailed fuzzy profile (Table 1) and a large number of documents were returned by the query (a total of 127 documents, see Figure 3).

The attempt to balance term-based and fuzzy searches (TF-IDF/Fuzzy) show that only in two of six searches at $k = 20$ did an improved ranking result compared to either TF-IDF or fuzzy. Both of these searches (Searches a and f) where the TF-IDF/fuzzy logic hybrid worked best were the most thematically non-specific (see Figure 3). Examining the specificity of the searches in Figure 3 and precision in Figure 4, it is not readily apparent that the raw number of hits has a direct relationship to the precision metric in general, or the performance of one ranking mechanism over another. Result independence with respect to specificity indicates that the ranking mechanisms are robust with respect to the size of the document set.

The computational complexity of TF-IDF and fuzzy systems are identical, and are dominated only by the step of vector-matrix multiplication. In each algorithm, the representation of the vector space can be created by repeated passes through the set of returned documents, using $O(n)$ time. The following step is to find the dot product of the search values and the document values, which is done in $O(n^2)$ time. Following evaluations, sorting will be at worst $O(n^2)$. The TF-IDF/fuzzy logic

Table 1 User profile attribute values set by expert

Search 1: geology Alberta
<Categories, {Geology & Geophysical = 1.0} >, <Locations, {Alberta = 1.0} >
Search 2: public infrastructure Nova Scotia
<Categories, {Public Health & Disease = 0.5} >, <Categories, {Public Safety & Security = 0.3} >, <Categories, {Transportation Networks = 0.8} >, <Categories, {Utilities & Communication = 1.0} >, <Location, {Nova Scotia = 1.0} >, <Location, {Atlantic Canada = 0.5} >
Search 3: road Ottawa
<Categories, {Transportation Networks = 1.0} >, <Locations, {Ontario = 1.0} >
Search 4: wildlife Labrador
<Categories, {Biology & Ecology = 0.5} >, <Categories, {Environment & Conservation = 0.7} >, <Location, {Newfoundland = 1.0} >, <Location, {Atlantic Canada = 0.5} >
Search 5: Canadian glaciers
<Categories, {Atmosphere & Climate = 0.5} >, <Categories, {Environment & Conservation = 0.5} >, <Categories, {Inland Water Resources = 0.5} >, <Locations, {Territorial North = 0.7} >, <Locations, {Innuition = 0.5} >, <Categories, {Alberta = 0.7} >, <Location, {British Columbia = 1.0} >, <Location, {Newfoundland = 0.5} >, <Location, {Northwest Territories = 0.5} >, <Location, {Nunavut = 0.5} >, <Location, {Yukon = 0.5} >
Search 6: water Canada
<Categories, {Inland Water = 1.0} >

hybrid is thus also dominated by $O(n^2)$. All algorithms are of reasonable time complexity for document ranking [4] and run in practical time.

6 Conclusions

In this work, we provide intelligent ranking of documents returned by a web portal that accesses multiple online Canadian-focused geographic data resources. To provide the ranking desired by an end user (the expert), we produced our own web portal prototype designed around a novel fuzzy logic ranking system for documents returned by the existing GeoConnections Discovery Portal. The precision of the top search results of five different ranking systems was examined: unranked, TF-IDF (an information retrieval baseline), fuzzy logic, and a TF-IDF and fuzzy logic hybrid solution. The expert user identified relevant documents using searches he devised to

test the capabilities of the system across both a spatial and thematic specificity spectra. The precision results indicated that the fuzzy logic system outperformed TF-IDF in almost all searches, so the geographic metadata in the returned set of items proved more valuable for effective ranking than the text content of the documents in this study. The fuzzy logic system generally provided increased effectiveness in ranking documents with spatial specificity at a moderate to high level. The fuzzy ranking system examined is also robust with respect to the number of documents in the result set. Future work may examine the possibility of improving the intelligent fuzzy logic search presented here by adding initial filtering on the GeoDiscovery Portal that uses spatial queries created by bounding box rectangles of quantitative latitude and longitudes and temporal queries.

Acknowledgements The authors would like to acknowledge programming assistance provided by Derek Leblanc, the financial assistance of a Canadian GEOIDE Network grant held by the second author, and the assistance of the GeoConnections initiative in accessing their web services.

References

1. R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Terminology Behind Search*. Addison-Wesley, 2nd edition, 2011.
2. S. Büttcher, C. Clarke, and G. Cormack. *Information Retrieval: Implementing and Evaluating Search Engines*. MIT Press, Cambridge, USA, 2010.
3. G. Castellano, D. Dell’Agnello, A. M. Fanelli, C. Mencar, and M. A. Torsello. A competitive learning strategy for adapting fuzzy user profiles. In *10th International Conference on Intelligent Systems Design and Applications, ISDA 2010*, pages 959–964, Nov. 29-Dec. 1 2010.
4. E.-H. Han and G. Karypis. Centroid-based document classification: Analysis and experimental results. In *Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, PKDD ’00*, pages 424–431, London, UK, 2000. Springer-Verlag.
5. M. Holi, E. Hyvnen, and P. Lindgren. Integrating tf-idf weighting with fuzzy view-based search. In *Proceedings of the ECAI Workshop on Text-Based Information Retrieval (TIR-06)*, 2006.
6. C. B. Jones and R. S. Purves. Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3):219–228, 2008.
7. M. A. A. Leite and I. L. Ricarte. Document retrieval using fuzzy related geographic ontologies. In *Proceeding of the 2nd International Workshop on Geographic Information Retrieval, GIR ’08*, pages 47–54, New York, NY, USA, 2008. ACM.
8. C. Mencar, M. Torsello, D. Dell’Agnello, G. Castellano, and C. Castiello. Modeling user preferences through adaptive fuzzy profiles. In *9th International Conference on Intelligent Systems Design and Applications, ISDA 2009*, pages 1031–1036, Nov. 30-Dec. 2 2009.
9. A. Micarelli, F. Gasparetti, F. Sciarrone, and S. Gauch. Personalized search on the world wide web. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, *The Adaptive Web: Methods and Strategies of Web Personalization*, pages 195–230. Springer-Verlag, 2007.
10. M. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.
11. T. Ross. *Fuzzy Logic with Engineering Applications*. John Wiley and Sons, Ltd., West Sussex, England, 2004.
12. N. O. Rubens. The application of fuzzy logic to the construction of the ranking function of information retrieval systems. *Computer Modelling and New Technologies*, 10(1):20–27, 2006.