Automatic Acquisition of Fuzzy Footprints

Steven Schockaert, Martine De Cock, and Etienne E. Kerre

Department of Applied Mathematics and Computer Science,
Fuzziness and Uncertainty Modelling Research Unit,
Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium
{Steven.Schockaert, Martine.DeCock, Etienne.Kerre}@UGent.be
http://www.fuzzy.ugent.be

Abstract. Gazetteer services are an important component in a wide variety of systems, including geographic search engines and question answering systems. Unfortunately, the footprints provided by gazetteers are often limited to a bounding box or even a centroid. Moreover, for a lot of non–political regions, detailed footprints are nonexistent since these regions tend to have gradual, rather than crisp, boundaries. In this paper we propose an automatic method to approximate the footprints of crisp, as well as imprecise, regions using statements on the web as a starting point. Due to the vague nature of some of these statements, the resulting footprints are represented as fuzzy sets.

1 Introduction

Information on the web is often only relevant w.r.t. a particular geographic context. To this end, geospatial search engines [5, 7] try to enhance the functionality of search engines by georeferencing web pages, i.e. by automatically assigning a geographic location to web pages. Consider for example the web page of a pizza restaurant in Gent. A geographic search engine would, for example, only return the web page of this restaurant if the user is located in East Flanders[1]. Geographic question answering systems [10] go even further, as they are able to respond to natural language questions and requests from users such as "What are the neighbouring countries of Belgium?". Clearly, these systems have to make use of some kind of digital gazetteer to obtain the necessary background knowledge. To respond appropriately to a request like "Show me a list of pizza restaurants in the Ardennes."[2] a suitable footprint of the Ardennes is needed. However for reasons discussed in [4], the footprints provided by gazetteers are often restricted to a bounding box, or even a point (expressed by its latitude and longitude coordinates). For imprecise regions such as the Ardennes, providing a bounding box is not even feasible since this kind of regions is not characterized by a clearly defined boundary.

A promising solution is to construct the footprint of a particular region in an automatic way. In [1] a method based on Voronoi diagrams for approximating

[1] Gent is the capital of the province of East Flanders.

[2] The Ardennes is a region in the southern part of Belgium.

R. Meersman et al. (Eds.): OTM Workshops 2005, LNCS 3762, pp. 1077–1086, 2005.

the footprint of a given region is proposed. Such a footprint is constructed from a set of points which are known to lie inside the region and a set of points which are known to lie outside the region. These sets of points are assumed to be correct and a priori available, e.g. provided by the user, hence this method is not fully automatic. In [9], it is suggested to represent regions with indeterminate boundaries by an upper and a lower approximation. Upper and lower approximations are constructed based on a set of points or regions which are known to be a part of the region under consideration, and a set of regions which are known to include the region under consideration. Again these sets are assumed to be correct and available a priori. A fully automatic algorithm is introduced in [2] where statements on the web such as "... in Luxembourg and other Ardennes towns ..." are used to obtain a set of points which are assumed to lie inside the region under consideration. To obtain a footprint from this set of points, the algorithm from [1] is slightly modified to cope with the noisiness of data from the web. Finally, in [8] kernel density surfaces are used to represent imprecise regions. However it is unclear what meaning should be attached to the weights corresponding to each point, as these weights seem to reflect the popularity (e.g. expressed as the number of occurrences of the corresponding city on the web) of the corresponding cities rather than some kind of vague representation of a region.

In this paper we introduce a new method to automatically construct a footprint for, possibly imprecise, regions by extracting relevant statements from the web. In contrast to existing approaches [2, 8] we do not only search for places that lie in the region under consideration, but also for regions that include this region, and for regions that are bordering on this region. Moreover, we use statements on the web such as "x is in the south-western corner of $\mathcal{R}$" to constrain the possible cities that could lie in the region $\mathcal{R}$. Due to the vagueness of this type of constraints, we propose using possibility distributions to this end. Information on the web can be inaccurate, outdated or even simply wrong. Hence, enforcing every constraint that is found on the web can result in an inconsistent solution (e.g. the only possible footprint is the empty region). Therefore, we apply ideas from the theory of fuzzy belief revision to (partially) discard certain constraints in the face of inconsistencies. The resulting footprint of the region is represented as a fuzzy set, which we call a fuzzy footprint in this context.

2 Obtaining Data from the Web

2.1 Acquiring Place Names Through Regular Expressions

Assume that we want to approximate the extent of a (possibly imprecise) region $\mathcal{R}$. The first step of our algorithm consists of searching the web for relevant statements and extracting useful data from these statements. In order to find relevant statements we send a number of queries to Altavista[3] such as "$\mathcal{R}$", "$\mathcal{R}$ is located in", "in $\mathcal{R}$ such as", ... and analyse the snippets that are returned.

[3] `http://www.altavista.com`

Table 1. Regular expressions

<table>
<tr><th colspan="2">Abbreviations</th></tr>
<tr><td><direction></td><td>= (heart | centre | ...| north-west | north-western)</td></tr>
<tr><td><place></td><td>= (village | villages | town | towns | city | cities)</td></tr>
<tr><td><area></td><td>= (region | province | state | territory)</td></tr>
<tr><td><names></td><td>= <name> (, <name>)* (and <name>)?</td></tr>
<tr><td><name></td><td>= [A-Z][a-z]+ ([A-Z][a-z]+)?</td></tr>
<tr><td><dir-part></td><td>= the? <direction> (part | corner)? of?</td></tr>
<tr><td><region-part></td><td>= the? ((<area> of)? R | R <area>?)</td></tr>
<tr><td><dir-reg></td><td>= <dir-part>? <region-part></td></tr>
</table>

<table>
<tr><th colspan="2">Regular expressions to find points inside $\mathcal{R}$</th></tr>
<tr><td>1.</td><td>(located|situated) in <dir-reg> (the <place> of)? <names></td></tr>
<tr><td>2.</td><td><names> (is|are) (a? <place>)? in <dir-reg></td></tr>
<tr><td>3.</td><td><names> (is|are) (located|situated) in <dir-reg></td></tr>
<tr><td>4.</td><td><names>, (located|situated) in <dir-reg></td></tr>
<tr><td>5.</td><td><names> and (a lot of)? other <place> in <dir-reg></td></tr>
<tr><td>6.</td><td><place> in <dir-reg> (are|such as|like|including) <names></td></tr>
</table>

<table>
<tr><th colspan="2">Regular expressions to find regions bordering on $\mathcal{R}$</th></tr>
<tr><td>7.</td><td><name> <area>? which borders the? (<area> of)? R</td></tr>
<tr><td>8.</td><td>R <area>? which borders the? (<area> of)? <name></td></tr>
<tr><td>9.</td><td><name> <area>? bordering (on|with)? the? (<area> of)? R</td></tr>
<tr><td>10.</td><td>R <area>? bordering (on|with)? the? (<area> of)? <name></td></tr>
</table>

Note that for reasons of efficiency we only analyse the snippets, and do not fetch the corresponding full documents. From these snippets we want to obtain:

1. A set P of points that are assumed to lie in $\mathcal{R}$.
2. The country $\mathcal{S}$ that is assumed to include $\mathcal{R}$[4].
3. A set B of regions that are assumed to border on $\mathcal{R}$.
4. A set C_P of constraints w.r.t. the positioning of some of the points in P (e.g. q is in the north of $\mathcal{R}$).
5. A set C_S of constraints w.r.t. the positioning of $\mathcal{R}$ in $\mathcal{S}$ (e.g. $\mathcal{R}$ is in the north of $\mathcal{S}$).

To this end we adopt a pattern-based approach using the regular expressions in Table 1. The regular expressions 1–6 can be used to find places in $\mathcal{R}$ and some corresponding constraints, i.e. to construct P and C_P. The regular expressions that are used to construct $\mathcal{S}$ and C_S (not shown) are entirely analogous. Finally, the regular expressions 7–10 can be used to find bordering regions, i.e. to construct B.

[4] If there are several possible countries found that may include $\mathcal{R}$, the algorithm could simply be repeated for each candidate, and the optimal solution could be selected afterwards. Furthermore, we could also consider the union of several (neighboring) countries to cope with regions whose extent spans more than one country.

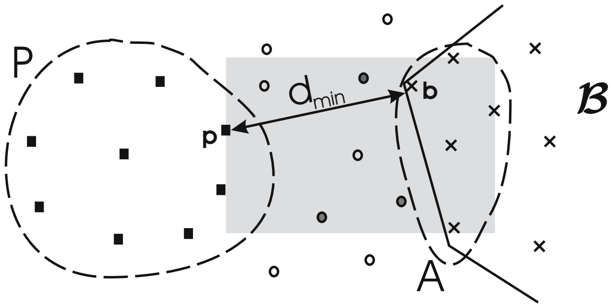

Fig. 1. Considering bordering region $\mathcal{B}$ to extend P. P is the set of points assumed to be in the region $\mathcal{R}$ to be approximated

2.2 Grounding the Place Names

The next step consists of grounding the geographic names that have been found. To accomplish this we use the Alexandria Digital Library (ADL) gazetteer[5], an online gazetteer service which can be accessed by using an XML- and HTTP-based protocol. To ground the place names in P, we first discard all names that are not located in $\mathcal{S}$. To disambiguate the remaining names, we choose the interpretations of the names in such a way that the area of the convex hull of the corresponding locations is minimal. This is a well-known heuristic which is, for example, described in more detail in [6]. To ground the bordering regions in B, we only consider administrative regions which are a part of $\mathcal{S}$, or border on $\mathcal{S}$. Since the footprint of most administrative regions provided by the ADL gazetteer is a point (the centroid of the region), for each bordering region $\mathcal{B}$ we construct a more accurate footprint by determining the convex hull of the places that are known to be in $\mathcal{B}$ by the ADL gazetteer. Consequently, for each bordering region $\mathcal{B}$ we calculate the minimal distance d_{min} from a point p in P to a place b in $\mathcal{B}$. Let A be the set of places in $\mathcal{B}$ for which the distance to p is less than $\lambda \cdot d_{min}$ where $\lambda \geq 1$. We now make the assumption that all places that lie within the minimal bounding box of $A \cup \{p\}$ and are not known to lie in the bordering region $\mathcal{B}$, lie in $\mathcal{R}$. Therefore, we add the most northern, most southern, most western and most eastern of these places to P. Note that one of these places will be p. Adding all places is not desirable, as this would influence too much the median of P and the average distance between the places in P. This process is illustrated in Figure 1.

3 Constructing Solutions

So far we have used information about the country to which $\mathcal{R}$ belongs as well as about bordering regions of $\mathcal{R}$ to update the set of points P that are assumed to lie in $\mathcal{R}$. In this section we use the remaining data retrieved from the web

[5] `http://www.alexandria.ucsb.edu/gazetteer/`

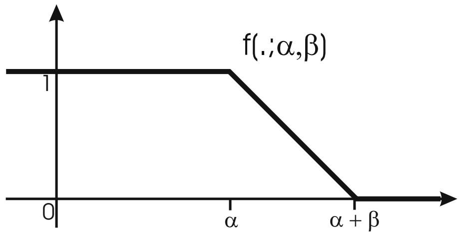

Fig. 2. The function $f(.;\alpha,\beta)$

to further enhance P, namely the sets of constraints C_P and C_S. Our aim is to construct a fuzzy set in P, i.e. a $P-[0,1]$ mapping F called a fuzzy footprint of $\mathcal{R}$. For each p in P, $F(p)$ is interpreted as the degree to which the point p belongs to $\mathcal{R}$. This membership degree is computed based on the constraints retrieved from the web. Each constraint is represented by a $P-[0,1]$ mapping c, called a possibility distribution in P. For every point p in P, $c(p)$ is the possibility that p lies in $\mathcal{R}$, taking into account the constraint modelled by c. For more information about fuzzy set theory and possibility theory, we refer to [12].

3.1 Modelling the Constraints

First, consider constraints of the form "q is in the north of $\mathcal{R}$", where q is a place in P. If p is south of q, the possibility that p lies in $\mathcal{R}$ remains 1. However, the further north of q that point p is situated, the less possible it becomes that p lies in $\mathcal{R}$. To construct the corresponding possibility distribution we use the function f depicted in Figure 2 as well as the average difference in y–coordinates between the points in P, i.e.

$$\Delta_y^{avg} = \frac{1}{|P|^2} \sum_{p\in P} \sum_{q\in P} |p_y - q_y| \tag{1}$$

The constraint "q is in the north of $\mathcal{R}$" is then modelled by the possibility distribution c_q^N, defined for each p in P as

$$c_q^N(p) = f(p_y - q_y; \alpha_1 \Delta_y^{avg}, \beta_1 \Delta_y^{avg}) \tag{2}$$

where $\alpha_1 \geq 0$ and $\beta_1 \geq 0$ are constants. Hence if $p_y - q_y \leq \alpha_1 \Delta_y^{avg}$ then the possibility that p lies in $\mathcal{R}$ is 1; if $p_y - q_y \geq (\alpha_1 + \beta_1)\Delta_y^{avg}$, then the possibility that p lies in $\mathcal{R}$ is 0; in between there is a gradual transition. In the same way, we can express that q is in the south, east or west of $\mathcal{R}$. Constraints of the form "q is in the north-west of $\mathcal{R}$" are separated in "q is in the north of $\mathcal{R}$" and "q is in the west of $\mathcal{R}$". Constraints of the form "q is in the middle of $\mathcal{R}$" can be represented in a similar way.

Constraints of the form "$\mathcal{R}$ is in the north of $\mathcal{S}$" are easier to model, since the exact bounding box of $\mathcal{S}$ is known. A point p from P is consistent with the constraint "$\mathcal{R}$ is in the north of $\mathcal{S}$" if it is in the northern half of this bounding box. Again, a fuzzy approach can be adopted where points that are slightly below the half are considered consistent to a certain degree. In the same way, we can express that $\mathcal{R}$ is in the south, east, west or centre of $\mathcal{S}$. Another type of constraints is induced by the elements of B. Each bordering region $\mathcal{B}$ in B induces a possibility distribution $c_{\mathcal{B}}$ on P, defined for each point p from P by $c_{\mathcal{B}}(p) = 0$ if p lies in $\mathcal{B}$ and $c_{\mathcal{B}}(p) = 1$ otherwise. In other words, if $\mathcal{B}$ is a bordering region of $\mathcal{R}$, $\mathcal{R}$ and $\mathcal{B}$ cannot overlap. In the following, let C_B be the set of all constraints induced by B.

Finally we impose an additional constraint c_h which is based on the heuristic that outliers in the set P are not likely to be correct. Let d be a distance metric on P (e.g. the Euclidean distance), we define the median m of P as $m = \arg\min_{p\in P} \sum_{q\in P} d(p,q)$. The possibility distribution c_h can be defined for each p in P by

$$c_h(p) = f(d(p,m); \alpha_2 d_{avg}, \beta_2 d_{avg}) \tag{3}$$

where $\alpha_2 \geq 0$ and $\beta_2 \geq 0$ are constants, and $d_{avg} = \frac{1}{|P|}\sum_{p\in P} d(m,p)$. In other words, the closer p is to the median m of P, the more possible it is that p lies in $\mathcal{R}$.

3.2 Resolving Inconsistencies

Let the set C be defined as $C = C_P \cup C_S \cup C_B \cup \{c_h\}$. Each of the possibility distributions in C restricts the possible places that could lie in $\mathcal{R}$. If each constraint were correct, we could represent the footprint of $\mathcal{R}$ as the fuzzy set F defined for p in P by

$$F(p) = \min_{c\in C} c(p) \tag{4}$$

This is a conservative approach in which the membership degree of p in $\mathcal{R}$ is determined by the constraint c that restricts the possibility of p lying in $\mathcal{R}$ the most. In practice however, C is likely to contain inconsistent information either because some websites contain erroneous information, because the use of regular expressions could lead to a wrong interpretation of a sentence, or because our interpretation of the constraints is too strict. As a consequence of these inconsistencies, F would not be a normalised fuzzy set, i.e. no point p would belong to F to degree 1, and could even be the empty set. To overcome this anomaly, we use a $C - [0,1]$ mapping K such that for c in C, $K(c)$ expresses our belief that c is correct. Formally, K is a fuzzy set in C, i.e. a fuzzy set of constraints. The fuzzy footprint corresponding with K is the fuzzy set F_K in P defined for p in P by

$$F_K(p) = \min_{c\in C} I_W(K(c), c(p)) \tag{5}$$

using the fuzzy logical implicator[6] I_W defined for a and b in $[0,1]$ by

$$I_W(a,b) = \min(1, 1-a+b) \tag{6}$$

[6] Implicators are $[0,1]^2 - [0,1]$ mappings which generalize the notion of implication from binary logic to the unit interval.

Eq. (5) expresses that we only impose the constraints in C to the degree that we believe they are correct. Note that if $K(c) = 1$ for all c in C (i.e. we are confident that all constraints are correct), then $F_K = F$. On the other hand if $K(c) = 0$ for all c in C (i.e. we reject all constraints), then $F_K = P$. The belief degrees in the constraints are determined automatically in a stepwise manner that can give rise to more than one optimal fuzzy set of constraints. We use $L^{(i)}$ to denote the class of optimal fuzzy sets A obtained in step i of the construction process; each A contains the first i constraints to a certain degree. Let $C = \{c_1, c_2, \ldots, c_n\}$ and $L^{(0)} = \{\emptyset\}$, i.e. $L^{(0)}$ is a set containing the empty set. For $i = 1, \ldots, n$ we define

$$L^{(i)} = \{A + c_i | A \in L^{(i-1)}\} \cup \{A \oplus c_i | A \in L^{(i-1)}\} \tag{7}$$

where $+$ is an expansion operator and $\oplus$ is a revision operator. The idea behind expansion is to add the next constraint c_i to A only to the degree α that c_i is consistent with A, i.e. to the highest degree α for which the footprint corresponding with the resulting fuzzy set of constraints is normalised. The idea behind revision is to select a particular fuzzy subset[7] $\hat{A}$ of A such that the footprint corresponding with $\hat{A}$ augmented with constraint c_i to degree 1 is normalised. In other words, for each constraint c_i that is not fully consistent with A we choose either to (partially) reject c_i, or to (partially) reject the constraints in A. For more details on fuzzy revision and expansion operators we refer to [3, 11].

If there are no inconsistencies, $L^{(n)}$ will contain only one fuzzy set K, hence F_K is the only possible footprint. However in the face of inconsistencies, $L^{(n)}$ will contain a number of possible alternatives $K_1, K_2, \ldots, K_m$. To rank the possible candidates, we assign each K_i a score $s(K_i)$ defined by

$$s(K_i) = \frac{area(cvx(F_{K_i}))}{\max_{j=1}^{m} area(cvx(F_{K_j}))} \cdot \frac{\sum_{c\in C} K_i(c)}{\max_{j=1}^{m} \sum_{c\in C} K_j(c)} \tag{8}$$

where for a fuzzy set B in $\mathbb{R}^2$

$$area(B) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} B(x, y) dx\ dy \tag{9}$$

provided the integral exists. The convex hull $cvx(B)$ of B is defined as the smallest convex fuzzy superset of B, i.e. every other convex fuzzy superset of B is also a fuzzy superset of $cvx(B)$, where a fuzzy set B in $\mathbb{R}^2$ is called convex if

$$\big(\forall\lambda \in [0, 1]\big)\big(\forall(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2\big)\big(B(\lambda x + (1 - \lambda)y) \geq \min(B(x), B(y))\big) \tag{10}$$

The score $s(K_i)$ expresses that optimal footprints should satisfy as many constraints as possible, while the corresponding extent should remain as large as possible.

---

[7] Let A and B be fuzzy sets in a universe X; A is called a fuzzy subset of B, or likewise B is a fuzzy superset of A, if and only if $(\forall x \in X)(A(x) \leq B(x))$.

4 Experimental Results

Because the footprint of an imprecise region is inherently subjective, we will focus in this section on political regions, which are characterized by an exact, unambiguous boundary. To this end, we will compare the fuzzy sets that result from our algorithm with a gold standard. As the gold standard for a region $\mathcal{R}$, we have used the convex hull of the place names that are known to lie in $\mathcal{R}$ by the ADL gazetteer. We will denote the gold standard for $\mathcal{R}$ by $\mathcal{R}^*$. Note that this gold standard is not a perfect footprint, among others because the *part-of* relation in the ADL gazetteer is not complete. Let $\mathcal{A}$ be a fuzzy set in $\mathbb{R}^2$; to assess to what extent $\mathcal{A}$ is a good approximation of $\mathcal{R}^*$, we propose the following measures:

$$s_p(\mathcal{A}) = incl(\mathcal{A}, \mathcal{R}^*) \qquad s_r(\mathcal{A}) = incl(\mathcal{R}^*, \mathcal{A})$$

where for A and B fuzzy sets[8] in a universe X

$$incl(A, B) = \frac{\sum_{x \in X} \min(A(x), B(x))}{\sum_{x \in X} A(x)} \tag{11}$$

s_p expresses the degree to which $\mathcal{A}$ is included in $\mathcal{R}^*$, i.e. the degree to which the places that lie in $\mathcal{A}$ also lie in $\mathcal{R}^*$; hence s_p can be regarded as a measure of precision. On the other hand, s_r expresses the degree to which $\mathcal{A}$ includes $\mathcal{R}^*$ and can be regarded as a measure of recall.

As test data we took 81 political subregions of France, Italy, Canada, Australia and China ("countries, 1st order divisions" in the ADL gazetteer). Table 2 and Table 3 show the values of $s_p(F_K)$ and $s_r(F_K)$ that were obtained using several variants of our algorithm, where F_K is the footprint with the highest score (Eq. (8)) that was constructed. As parameter values, we used $\alpha_1 = 0.5$, $\beta_1 = 1$, $\alpha_2 = 1.5$, $\beta_2 = 5$ and $\lambda = 1.5$. For the first four columns we didn't consider bordering regions (neither to extend P as in Section 2.2 nor to construct a set of constraints C_B); for the column 'no' no constraints were imposed, for 'c_h' only c_h was imposed, for 'C_P' only the constraints in C_P were imposed, and finally for 'C_P, C_S, c_h' the constraints in $\{c_h\} \cup C_P \cup C_S$ were imposed. For the last four columns, bordering regions were used to extend P as in Section 2.2. For the column 'all' the constraints in $\{c_h\} \cup C_P \cup C_S \cup C_B$ were imposed. Obviously for popular regions we will find more relevant cities, constraints and bordering regions. Therefore, we split the regions into three groups: regions for which at least 30 possible cities were found (11 regions), regions for which less than 10 possible cities were found (38 regions), and the other regions (32 regions). For popular regions, imposing the constraints significantly increases precision. Furthermore considering bordering regions significantly improves recall, provided not all constraints are imposed. Unfortunately, considering bordering regions also decreases precision drastically. We believe that this is, at least partially, caused by the fact

[8] Note that $\mathcal{R}^*$ is in fact an ordinary set. However ordinary sets can be treated as special cases of fuzzy sets for which the membership degrees take only values in $\{0, 1\}$.

Table 2. Precision $s_p(F_K)$

	no bordering regions				bordering regions			
	no	c_h	C_P	C_P, C_S, c_h	no	c_h	C_P	all
All regions	0.26	0.43	0.43	0.47	0.16	0.30	0.35	0.42
$\|P\| \geq 30$	0.35	0.70	**0.85**	**0.83**	0.15	0.43	0.57	0.62
$10 \leq \|P\| < 30$	0.31	0.51	0.51	0.57	0.19	0.36	0.43	0.51
$\|P\| < 10$	0.20	0.28	0.23	0.28	0.14	0.22	0.22	0.28

Table 3. Recall $s_r(F_K)$

	no bordering regions				bordering regions			
	no	c_h	C_P	C_P, C_S, c_h	no	c_h	C_P	all
All regions	0.49	0.39	0.35	0.32	0.57	0.49	0.44	0.37
$\|P\| \geq 30$	0.85	0.59	0.38	0.33	**0.91**	0.70	0.55	0.39
$10 \leq \|P\| < 30$	0.68	0.56	0.50	0.48	0.75	0.66	0.57	0.52
$\|P\| < 10$	0.23	0.19	0.21	0.17	0.32	0.28	0.30	0.25

that the *part-of* relation in the ADL gazetteer is not complete. Therefore if recall is considered less important than precision, bordering regions should not be used. On the other hand, if recall is considered more important than precision bordering regions should be used, but not all constraints should be imposed, e.g. only c_h or only the constraints in C_P.

5 Conclusions

We have proposed a novel method to approximate the footprint of a (possibly imprecise) region by using statements on the web as a starting point. Existing approaches consider only statements that express that a particular city lies in the region of interest. We have extended this by also considering bordering regions and regions that are assumed to include the region of interest. Moreover, we have proposed to interpret vague restrictions such as "x is in the north-western corner of $\mathcal{R}$" and thus reducing the noise which is inevitably apparent when using data from the web. As a consequence, the resulting footprint is represented as a fuzzy set instead of, for example, a polygon. Inconsistencies between the constraints are resolved by using ideas from the theory of (fuzzy) belief revision. The experimental results show that imposing constraints can significantly improve precision, while considering bordering regions improves recall.

Acknowledgments

Steven Schockaert and Martine De Cock would like to thank the Fund for Scientific Research – Flanders for funding their research.

References

1. Alani, H., Jones, C.B., Tudhope, D.: Voronoi-based region approximation for geographical information retrieval with gazetteers. Int. J. Geographical Information Science **15** (2001) 287–306
2. Arampatzis, A., van Kreveld, M., Reinbacher, I., Jones, C.B., Vaid, S., Clough, P., Joho, H., Sanderson, M., Benkert, M., Wolff A.: Web-based delineation of imprecise regions. Proc. of the Workshop on Geographic Information Retrieval, SIGIR. http://www.geo.unizh.ch/~rsp/gir/ (2004)
3. Booth, R., Richter, E.: On revising fuzzy belief bases. UAI International Conference on Uncertainty in Artificial Intelligence (2003) 81-88
4. Hill, L.L.: Core elements of digital gazetteers: placenames, categories, and footprints. Lecture Notes in Computer Science **1923** (2000) 280–290
5. Jones, C.B., Purves, R., Ruas, A., Sanderson, M., Sester, M., van Kreveld, M., Weibel R.: Spatial information retrieval and geographical ontologies: an overview of the SPIRIT project. Proc. of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2002) 387–388
6. Leidner, J.L., Sinclair, G., Webber, B.: Grounding spatial named entities for information extraction and question answering. Proc. of the Workshop on the Analysis of Geographic References, NAACL-HLT (2003) 31–38
7. Markowetz, A., Brinkhoff, T., Seeger, B.: Geographic information retrieval. Proc. of the 3rd International Workshop on Web Dynamics. http://www.dcs.bbk.ac.uk/webDyn3/webdyn3_proceedings.pdf (2004)
8. Purves, R., Clough, P., Joho, H.: Identifying imprecise regions for geographic information retrieval using the web. Proc. of the GIS Research UK 13th Annual Conference (to appear)
9. Vögele, T., Schlieder, C., Visser, U.: Intuitive modelling of place name regions for spatial information retrieval. Lecture Notes in Computer Science **2825** (2003) 239–252
10. Waldinger, R., Appelt, D.E., Fry, J., Israel, D.J., Jarvis, P., Martin, D., Riehemann, S., Stickel, M.E. , Tyson, M., Hobbs, J., and Dungan, J.L.: Deductive question answering from multiple resources. In: Maybury, M. (ed.): New Directions in Question Answering. AAAI Press (2004) 253–262
11. Witte, R.: Fuzzy belief revision. 9th Int. Workshop on Non-Monotonic Reasoning (2002) 311–320
12. Zadeh, L.A.: Fuzzy sets as the basis for a theory of possibility. Fuzzy Sets and Systems **1** (1978) 3–28