Oh Web Image, Where Art Thou?

Dirk Ahlers[1] and Susanne Boll[2]

[1] OFFIS Institute for Information Technology, Oldenburg, Germany
ahlers@offis.de
[2] University of Oldenburg, Germany
boll@informatik.uni-oldenburg.de

Abstract. Web image search today is mostly keyword-based and explores the content surrounding the image. Searching for images related to a certain location quickly shows that Web images typically do not reveal their explicit relation to an actual geographic position. The geographic semantics of Web images are either not available at all or hidden somewhere within the the Web pages' content. Our spatial search engine crawls and identifies Web pages with a spatial relationship. Analysing location-related Web pages, we identify photographs based on content-based image analysis as well as image context. Following the photograph classification, a location-relevance classification process evaluates image context and content against the previously identified address. The results of our experiments show that our approach is a viable method for Web image location assessment. Thus, a large number of potentially geographically-related Web images are unlocked for commercially relevant spatial Web image search.

1 Introduction

The relation of Web information to a physical place has gained much attention in recent years both in research and commercial applications. Web search started to become "local", meaning that users can search and access information that is mapped to an exact place or region of interest. However, currently a Web search for photos of a certain place remains at a keyword-level of known geographic names such as "New York" or "Singapore" but does not deliver a precise georeference for Web media content. To enable users to search for images at a specific geographical coordinate or its spatial vicinity, we need a Web image search with geo-spatial awareness.

The Web pages we find today seldomly provide explicit location-information. This also applies to the images contained in Web pages. Even though GPS has arrived at the end consumer, photos with an actual location stamp cover only a tiny fraction of all Web images. The challenge that remains for Web image retrieval is the derivation of reliable location information for Web images from unstructured Web pages. In our approach to spatially index Web images, we exploit address information that is hidden in Web pages. Employing our spatial Web search engine, we crawl and index Web pages for location information. If our geoparser identifies a reliable relaion between a Web pages' content and a

S. Satoh, F. Nack, and M. Etoh (Eds.): MMM 2008, LNCS 4903, pp. 101–112, 2007.

physical address, we can then determine whether the location of the Web page itself can also be assigned to the images embedded in it. Analysing the image content, we distinguish photographs from other images. Exploiting image and surrounding page content, a multi-criteria classification determines if the photograph can actually be related to the location identified for the Web page. We combine several heuristic approaches for analyzing and scoring relevant features to achieve an overall assessment of an image's connection to a location. Our test and evaluation show that only a small fraction of Web pages contain location-relevant photos but that these can be very well related to the spatial context. With a spatial Web image search engine thus established, we can answer the question "Oh Web image, where art thou?" Our spatial Web image index can then be applied for a variety of interesting applications such as searching for images along a route, extending personal photo collections, or illustrating and enhancing search results.

The paper is structured as follows. We present the related work in Section 2. Section 3 introduces the reader to our spatial search engine. In Section 4, we present our location-based image search approach. We evaluate and discuss our results in Section 5 and conclude the paper in Section 6 and give a perspective on ongoing work.

2 Related Work

One of the related fields is the *extraction of geographic meaning from unstructured Web pages* (Web-Geo-IR). An comprehensive overview of the state-of-the-art in this large field is presented in [8]. Central challenges in the field are the identification of geographical entities as described in [14] and [15] and the use of geographic features within a Web crawler as, e.g., discussed in [6]. The basis for our spatial Web image search is a spatial search engine we developed that employs geographically focused Web crawling and database-backed geoparsing to spatially index Web pages as presented in [2,1].

Related approaches for *Web image classification* can be grouped by their source of information, i.e., as HTML code and structure and Web image content. As an example for solely using the HTML code, [22] examines the attributes of the image tag, surrounding paragraph and title of the page. Aiming to understand and derive roles of Web images, [13] uses features taken from structural document data and image metadata but not from actual image content. Analysing HTML structure, [7] segments Web pages into semantic blocks whose semantics are inherited by embedded images. An early work [5] uses the HTML content and a few selected image content features for Web image search. Building upon this work, [3] developed a classifier for photographs against graphics using decision trees. The most recent work presents iFind [9], which considers context and content of images and examines their layout on a page for semantic annotation. A comprehensive discussion of existing technologies in Web image retrieval and how they address key issues in the field can be found in [11].

Considering the large field of *content-based image retrieval*, a good overview of the existing work, the state-of-the-art and challenges in the field can be found in [20] and [12]. Even though Web image analysis is not the explicit focus of many of these works, it became clear in the field of content-based retrieval that results in image understanding can only be advanced by looking at both the content and external sources of information.

Regarding the connection of location and Web images, some work has been done on deriving location by using only image content analysis for known locations. IDeixis [21] is an approach to compare pictures taken by a mobile device to previously taken photos to initiate a search. Work about inferring semantics by multimodal analysis of images and their spatial and temporal properties was done by, e.g., our group in [19]. On the way to location-aware images, Web 2.0 services that allow organizing and sharing personal photo collections provide for an assignment of location-related information by manual tagging and annotation. Often, user-chosen tags include names of locations or places. Some services such as Mappr (mappr.com), Flickr (flickr.com) or Placeopedia (placeopedia.com) even allow for assigning items a definite coordinate on a map and thus manually geocode the content. Content from existing directories such as yellow pages is mapped by services such as Yahoo! Maps (maps.yahoo.com) or MSN Live (maps.live.com), enabling map-based spatial search. Recently, Google Maps (maps.google.com) has allowed business owners to enhance entries with images of their business, clearly emphasizing the demand for accompanying images to location information.

Modern cameras with an integrated GPS-receiver can directly embed coordinated in a photo's EXIF metadata. However, such geo-located images are rare in the first place even in large photo collections and are even rarer in a general Web context, partly due to loss of information in editing processes. Regarding loss of metadata, [16] presents a system that captures metadata at media capture time and preserves it over the production process in a professional environment, but which ios not suitable for common use.

Unfortunately, tagged media collections or prepared geocoded images cover only a very small fraction of images on Web pages; many others still reside unrecognised in the World Wide Web. The remaining open issue is that explicit location information for Web pages and their images is given only for a small fraction of the Web which leaves a large number of potentially geographically-related images on Web pages as yet unrecognised for spatial search. We therefore need other methods of location assessment and additional sources of data for geo-referencing Web images.

3 Geo-spatial Search

Most current search engines are very efficient for keyword-based queries. Spatial search, i.e., querying for Web pages related to a certain location is starting to gain momentum. Services operating on prepared geo-referenced data receive widespread attention. This is consistent with the high relevance of location to the user. A study in 2004 [18] finds that as much as 20% of user's Web queries have a geographic relation while [10] estimates this at 5-15% for search on mobile devices. Web pages with location-related information represent an estimate of 10-20% of all Web pages. This makes their relation to a physical location a yet widely unused asset. A spatial Web search engine can discover and process this information for an interesting set of location-based applications.

For our own spatial Web search engine presented in [2], we use common Web crawler and indexing technology and extended it with our own location specific components. Our spatial crawler and indexer identify Web pages that promise a relation to a physical location, extract the desired information, and assign a geo-reference to the pages. They are designed to exploit the location information of a Web page that is not explicitly contained as metadata or structured annotation but is rather an implicit part of the textual content. Based on our research in mobile pedestrian applications [4], we tailor our search to a pedestrian user who needs precise geo-references at the level of individual buildings.

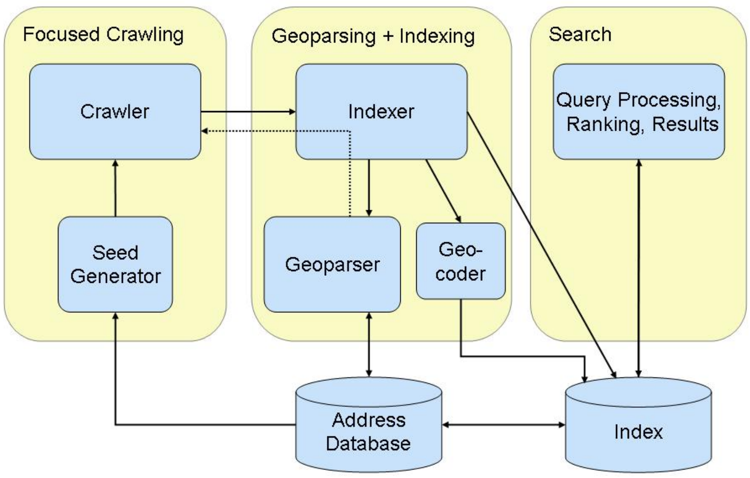


Fig. 1. Architecture of our spatial search engine

The architecture detailed in Figure 1 comprises the main components of our spatial Web search engine prototype, separated into function groups of crawling, indexing, and search. To enable an efficient geographically focused crawling, a seed generator feeds seed URLs that have a strong relationship to the desired physical location—such as, e.g., municipal sites—to enable the crawler to start with relevant pages. The crawler hands the pages found over to the indexer. Besides the common analysis of the textual Web content, a geoparser processes the page to extract geographic information. The extraction process is supported by an address database for enhanced accuracy. The key aspect of focused crawling is a feedback loop based on a page classifier to guide the crawler towards pages with the desired topic. Thus, the geoparser results are fed back to the crawler. Discovered location-related information is geo-coded, i.e., the hierarchical textual description is mapped to a coordinate. Along with the actual address and the URL it is inserted into a spatial index. The index of our search engine can then be used for a keyword-based search now extended by spatial search capability.

4 Web Image Location-Detection

While today an interesting set of semantics can be derived from image content, geographic location is usually not part of it. Apart from well-known geographic landmarks or existing annotations Web images cannot be processed by current techniques to assess and pinpoint an actual location.

The goal of our spatial Web image search is to be able to automatically illustrate locations with relevant images of their surroundings and for a given place to answer questions like "What does this place look like?" or "What is a representative image of this place?". We aim at a general approach to derive the location for those images that are contained in Web pages with an established spatial context. The mutual relation of a Web pages' text, its contained images, and accociated metadata is a main aspect of this method: The page's spatial semantics are extended to the embedded images which then will, to a certain degree and depending on detected features, inherit the location of the page.

Fig. 2. Images and location on a sample Web page

As one example of our approach, consider the Web page of a public library in Germany as shown in Figure 2. The page features various images – surrounded by dashed lines – with the one in the center prominently displaying the library building. Its address is present to the right of the image and its name and city in some of the metadata such as keyword and title. For a location-based image search, only this photograph should be retrieved since it is descriptive of the location.

4.1 Architecture of the Location Assessment

For the identification of a possible location-relation of images from Web pages, we propose a method and architecture for image location assessment. Using only the Web page data is not sufficient as only preliminary conclusions can be drawn about the image content. Thus, this has to be taken into account as well. The main logical components are two classifiers for photographs and for location relevance, respectively. These are arranged as weighted filter chains and utilize various criteria taken from the HTML content and metadata, the metadata of embedded images, image content, and the address previously found. The general process is outlined in Figure 3. Input is a Web page with a known location from, e.g., our spatial search. All images are extracted, downloaded and processed by a photograph classification, followed by a location assessment. Image and Web page are examined for hints towards a location reference, which are weighted and summarised. If a location can be reliably assigned, the process results in a geo-coded image. In the following we describe the two classifications.

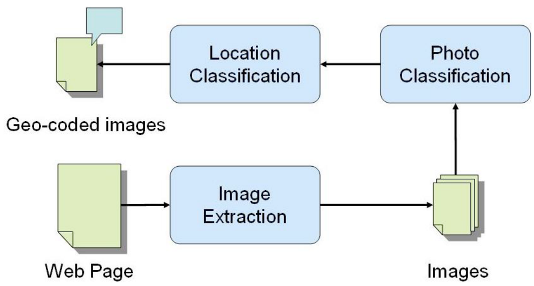


Fig. 3. Process of the location assessment

Photograph classification. Today, images on Web pages are used for different reasons. Distinguishing their different roles supports filtering certain images by their roles. A large group of images are those that serve an *illustrative* purpose: photographs, diagrams, graphics, sketches, titlebars, logos. Other images may serve as *formatting objects*, represent *advertisements* or are used for *user tracking*. Of these roles, we are interested in *photographs* being the main bearers of visual location information. Such photographs serve illustrative purposes; page authors use them with this in mind, and they clearly depict objects or scenery that is in most cases related to the text they accompany. This corresponds to the results of our initial surveys and probings. Therefore, we developed methods to identify images with a high probability to be a photograph of a place, drawing from low-level image features. While each of these features are inconclusive, in combination they yield satisfying results.

- *Image size and image ratio* is used to remove small formatting images as well as illustrative images that lack sufficient size or that have an unusual ratio such as banners etc.

- *Color count and histogram analysis* is used to distinguish photographs from illustrations, drawings or logos. Black & white images are detected to be treated with different thresholds.
- *Animation or transparency* in images usually indicates non-photographs.
- *EXIF-data* in JPG files is examined. Specific metadata such as camera type, focal length, flash usage etc. is considered strong evidence that an image was taken by a digital camera and is thus a photograph.
- *Vertical edges* in above-average quantity can indicate photographs of a place showing buildings, pieces of architecture or trees. For our purposes, a modified Sobel process operating on greyscaled images for edge detection worked best for urban buildings but is inconclusive for rural scenery.
- *Face detection* is used to distinguish mere snapshots of people at a location from images truly depicting the location itself since we found the number of such snapshots quite high.

Location-relevance classification. An image related to a location will exhibit certain characteristics which we exploit to create a set of measurements of how well the location of a Web page can be inherited by the images embedded in it. To establish a relation of a photograph and the location referenced on the Web page, we rely on content and structure of the page in question. From the photograph classification, we know all photographs on the page; from our spatial search engine, we know the geographic location of a Web page as a full address including street, number, zip code, city name and also geo-coordinate.

However, we assume this location reference may not necessarily stand for the entire page, it may rather determine a location for only a part of the page. We assume a decreasing relevance by increasing distance of the address and the image on the basis of a DOM-tree representation of the page. An image in the vicinity of an address has a higher probability for location relevance than one further away; if we can identify a page as generally dealing with the location, the probability also increases.

For the implementation of the location-relevance classification, we consider the distance of the image to each address on the page and the matching of image descriptors and key elements on the page to parts of the address. We use an in-page location keyword search. We interpret the given address as a small subsection of a geographic thesaurus. This allows for generalization of an address by traversing its hierarchy for keywords. We can broaden the search from full street address to only city or region. An address such as "*Escherweg 2, 26121 Oldenburg*" could be matched in decreasing levels of relevance as "*Escherweg, Oldenburg*" or "*Oldenburg*" alone. This truncation matching is already useful for searching the page, but absolutely necessary for searching features such as metatags, page title, or image attributes. They are very unlikely to contain full addresses, but will often contain at least the city name. Matches of this search are rated by the amount of matched keywords and—where applicable—the distance from the image based on the DOM-representation of the document. The process would repeat for each location associated to a Web page.

The examined image tag attributes are *alt*, *title*, and *name*. These descriptive fields for an image may contain descriptions of the image contents and are therefore highly relevant as well as the *src* attribute for meaningful image or path names. The page's metatag fields *description*, *keywords* and Dublin Core *DC.Subject*, *DC.Description* as well as page title can hint that the page as a whole is concerned with the location and therefore the parts on it get a higher rating. Other well-known tags such as dc.coverage, geo.location or icbm coordinates were not encountered during our crawls and were therefore not included at this time.

4.2 Combining Features for Location Assessment

The process in Figure 3 depicts two logical blocks for location assessment, photograph classification and location classification, each realised by an analysis chain. Inside each chain, several modules evaluate the aforementioned features and calculate a score for the assessed image.

The assessment chains are defined as follows: Let n be the number of evaluation modules in the chain. Each module $evaluation_i$ is assigned a weight w_i to express its relevance relative to other modules and allow for easy parameterization and tuning. The relevance score for an image img is calculated by summing up the evaluation scores for each feature:

$$relevance(img)_{chain} = \sum_{i=1}^{n} w_i \; evaluation_i(img) \tag{1}$$

For each chain, the final relevance score of an image is compared to an empirically determined threshold $threshold_{chain}$ to decide whether it is in or out of scope.

$$pass(img)_{chain} = \begin{cases} 1 & threshold_{chain} \leq relevance(img)_{chain} \\ 0 & otherwise \end{cases} \tag{2}$$

Having defined the general evaluation chains, we install two actual chains *photo* and *location* for photograph detection and location assessment respectively. An image is classified as location-relevant if it passes both chains; its overall location score is given as $relevance(img)_{location}$ and the relation between image and location is established.

5 Evaluation

To evaluate our approach, we performed a crawl of Rügen, Germany's largest island and a famous tourist area, and of the city of Oldenburg, Germany. The resulting testset comprises Web pages along with the addresses that were identified on them. In the following, we discuss the results of the crawl of Rügen only as the second crawl of the city of Oldenburg delivered quite similar results. We show the performance of photograph classification on a smaller testset and then present the location classification in a second step.

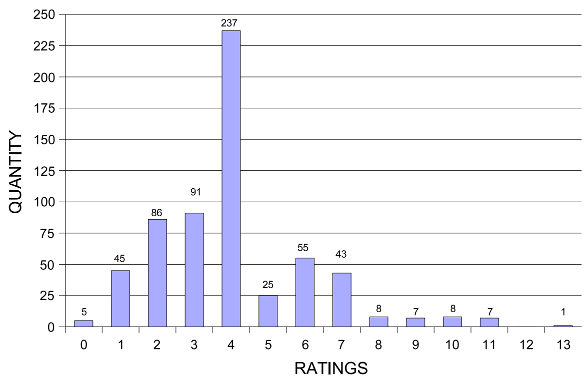
5.1 Evaluation of Photograph Classification

To create a small testset for the photograph classification we used a random sampling from our spatial index of 100 Web pages with an identified location relevance. Out of these, 78 pages could be used as 22 pages remained unreachable or had parsing problems. Within the remaining pages, we found 1580 images. Our photograph classification resulted in 86 images classified as photographs. Visual inspection of these images revealed that 69 were classified correctly. The remaining 17 images were misclassified graphics which still showed some of the relevant photographic features. 1538 images were classified as non-photographs and mainly comprised decorative items of Web pages. We also went through a visual inspection of these images and identified 1511 as classified correctly. The 27 falsely classified photographs contained only few colors or had untypical sizes or ratios such that their score fell beyond the threshold. The fact that we only found very few photographs is not surprising, as on the Web nowadays most images are used for decorative purposes. Within this small testset, the achieved values for $precision \approx 0,80$ and $recall \approx 0,72$ are promising.

5.2 Evaluation of Location Classification

To draw meaningful conclusions for the location assessment, we enlarged the testset to 1000 pages. Of these, 857 pages could be examined. A glance on the resulting photographs confirmed a similar performance as before, but was not manually assessed. A resulting 618 combinations of photos with locations were analysed. The ratings are depicted in Figure 4 as well as one sample image with a score of 13. An accumulation of images at the lower end of the scale around 4 can be observed. Since an address was present on the page, some evaluators such as city name matching generate a low initial score. But only with the combination of several well-matching criteria an image's score raises sufficiently to reach a high ranking above the determined threshold of 6 in our system. Manual inspection shows that indeed this group contains highly relevant images while low-rated images have mostly no discernible location associated.

Of the 74 images classified as having a location relevance we found 55 classified correctly and 19 with dubious or false results. Of the 544 images classified without location relevance, we have no exact values for the ratio of wrong classification. However, scanning these images seems to confirm that most of them are correctly classified; we estimate recall at above 0.7 and can calculate $precision \approx 0,74$. Examples of the images retrieved by our system are shown in Fig. 5. We find some well-related images *a*-*d*, some even from the interior of buildings, especially for holiday homes. However, we also return some false positives and mismatches. Images *e* and *g* depict a location, but the connection to the found location is uncertain. For *g* we can assume that it is just a general landscape of the island. Some uncertainties could only be solved by in-depth linguistic analysis of the pages, some are unsolvable even by a human user. Clearly mis-classified is the illustrative graphic *f*. The sketch of a chapel *h* shows good location relevance but is not a photo. It shows an interesting capability of our system: The major part of images falsely



(a)

(b)

Fig. 4. (a) Distribution of location ratings (b) Example of the top-rated image

Fig. 5. Retrieved images, sample true positives (a–d) and false positives (e–h)

classified as photographs were rated lower for location relevance than true photographs. Apparently the location classification compensates for some errors in the photograph detection.

6 Conclusion

In this paper, we presented an approach to automatically geo-reference images embedded in Web pages on the address level. Based on our work on location-based Web search, we presented our method to reliably determine photographs on Web pages and the heuristics to determine if the location of the Web page can be reliably assigned to the photograph. We could show that a combined analysis of content and context is feasible for this task. Automatic location detection of images on unstructured Web pages now allows images to become part of the results of a spatial Web image search without previous manual tagging.

The initial results of our experiments are promising and show that we can achieve a precision for the location classification of about 0.74 with an estimated recall of 0.7. As the discussion of the evaluation revealed, it remains challenging to determine if a photo carries a location-relation. Still, the results and our chosen features show that context analysis for images alone is not sufficient but in combination with basic image content analysis we can reliably detect location-bearing photographs. In our ongoing work we will refine the parameters of the heuristics, analyse in more detail to what extent the filters each influence the results and how this can be exploited in more sophisticated classification by, e.g., decision trees and the use of more features for the classification. We will further increase the quality of our Web image location search.

Acknowledgements

We thank Jörg Baldzer and Norbert Rump for their work on the search engine project as well as our students Sebastian Stockfleth, Timo Scheffler, and Dorothea Eggers. Part of this research has been carried out as subproject of the Niccimon project [17] and was supported by the State of Lower Saxony, Germany.

References

1. Ahlers, D., Boll, S.: Geospatially Focused Web Crawling. Datenbankspektrum, Special Issue Focused Search (Fokussierte Suche) (2007)
2. Ahlers, D., Boll, S.: Location-based Web search. In: Scharl, A., Tochtermann, K. (eds.) The Geospatial Web, Springer, Heidelberg (2007)
3. Athitsos, V., Swain, M.J., Frankel, C.: Distinguishing photographs and graphics on the World Wide Web. In: CAIVL 1997: Proceedings of the 1997 Workshop on Content-Based Access of Image and Video Libraries, p. 10. IEEE Computer Society, Los Alamitos (1997)
4. Baldzer, J., Boll, S., Klante, P., Krösche, J., Meyer, J., Rump, N., Scherp, A., Appelrath, H.-J.: Location-Aware Mobile Multimedia Applications on the Niccimon Platform. In: IMA 2004 (2004)
5. Frankel, C., Swain, M.J., Athitsos, V.: WebSeer: An Image Search Engine for the World Wide Web. Technical report, Chicago, IL, USA (1996)
6. Gao, W., Lee, H.C., Miao, Y.: Geographically focused collaborative crawling. In: WWW 2006: Proc. of the 15th Intl. Conf. on World Wide Web, pp. 287–296. ACM Press, New York (2006)
7. Gong, L.H.U.Z., Cheang, C.W.: Web image semantic clustering. In: ODBASE 2005. Proceedings of the 4th International Conference on Ontologies, Database and Applications of Semantics. LNCS, Springer, Heidelberg (2005)
8. Gräf, J., Henrich, A., Lüdecke, V., Schlieder, C.: Geografisches Information Retrieval. Datenbank-Spektrum 6(18), 48–56 (2006)
9. He, X., Caia, D., Wen, J.-R., Ma, W.-Y., Zhang, H.-J.: Clustering and searching WWW images using link and page layout analysis. ACM Trans. on Multimedia Comput. Commun. Appl. 3 (2007)

10. Kamvar, M., Baluja, S.: A large scale study of wireless search behavior: Google mobile search. In: CHI '06: Proc. of the SIGCHI conf. on Human Factors in computing systems, pp. 701–709. ACM Press, New York (2006)
11. Kherfi, M.L., Ziou, D., Bernardi, A.: Image retrieval from the world wide web: Issues, techniques, and systems. ACM Comput. Surv. 36(1), 35–67 (2004)
12. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimedia Comput. Commun. Appl. 2(1), 1–19 (2006)
13. Maekawa, T., Hara, T., Nishio, S.: Image classification for mobile web browsing. In: WWW 2006: Proc.of the 15th Intl. Conf. on World Wide Web, pp. 43–52. ACM Press, New York (2006)
14. Markowetz, A., Chen, Y.-Y., Suel, T., Long, X., Seeger, B.: Design and implementation of a geographic search engine. In: WebDB 2005, Baltimore (2005)
15. McCurley, K.S.: Geospatial mapping and navigation of the web. In: WWW 2001: Proc. of the 10th Intl. Conf. on World Wide Web, pp. 221–229. ACM Press, New York (2001)
16. Nack, F., Putz, W.: Designing annotation before it's needed. In: MULTIMEDIA 2001: Proceedings of the ninth ACM international conference on Multimedia, pp. 251–260. ACM Press, New York (2001)
17. Niccimon. Competence center of lower saxony for information system for mobile usage (2006), `http://www.niccimon.de/`.
18. Sanderson, M., Kohler, J.: Analyzing geographic queries. In: Workshop on Geographic Information Retrieval SIGIR (2004)
19. Sandhaus, P., Scherp, A., Thieme, S., Boll, S.: Metaxa – context- and content-driven metadata enhancement for personal photo albums. Multimedia Modeling 2007, Singapore (2007)
20. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. 22(12), 1349–1380 (2000)
21. Tollmar, K., Yeh, T., Darrell, T.: Ideixis - image-based deixis for finding location-based information. In: MobileHCI (2004)
22. Tsymbalenko, Y., Munson, E.V.: Using HTML Metadata to Find Relevant Images on the World Wide Web. In: Proceedings of Internet Computing 2001, Las Vegas, vol. 2, pp. 842–848. CSREA Press (2001)