Local Web Search Examined

Dirk Ahlers

Unitec – Universidad Tecnológica Centroamericana

Sistemas Computationales, Facultad de Ingeniería

Tegucigalpa, Honduras

ahlers@dhere.de

Purpose

To provide a theoretical background to understanding current local search engines as an aspect of specialized search, and understanding the data sources and used technologies.

Design/Methodology/Approach

Selected local search engines are examined and compared towards their use of GIR (Geographic Information Retrieval) technologies, data sources, available entity information, processing and interfaces. An introduction to the field of GIR is given and its use in the selected systems is discussed.

Findings

All selected commercial local search engines utilize GIR technology in varying degrees for information preparation and presentation. It is also starting to be used in regular Web search. However, major differences can be found between the different search engines.

Research limitations/implications

This study is not exhaustive and only uses informal comparisons without definitive ranking. Due to the unavailability of hard data, informed guesses were made based on available public interfaces and literature.

Practical implications

A source of background information for understanding the results of local search engines, their provenance and their potential.

Originality/value

An overview of GIR technology in the context of commercial search engines integrates research efforts and commercial systems and helps to understand both sides better.

Keywords: Local Search; Web Search; Geospatial Search; Geographic Information Retrieval;

Location-based Services.

Papertype: General Review

The final version is published by Emerald as
Dirk Ahlers (2012), Chapter 3 Local Web Search Examined, in Dirk Lewandowski (ed.) Web Search Engine Research (Library and Information Science, Volume 4) Emerald Group Publishing Limited, pp.47 - 78
http://dx.doi.org/10.1108/S1876-0562(2012)002012a005

1. Introduction

Local search today is a major and popular form of specialized Web search. For many geographical information needs, it allows to easily and intuitively access information about specific places, focus search requests to specific regions, or provide pinpointed information on a user's whereabouts (Ahlers & Boll, 2007). With the ubiquity of mobile devices, the standard search engine paradigm of sending a request and receiving a list of matching web pages is complemented by new and interesting search applications with different interaction strategies, both mobile and for the desktop. However, depending on the information need, different aspects of local search and localizable data need to be understood.

Should the search result be a list of Web pages dealing with a geographical aspect or describing a place, photos showing a landmark, nearby validated businesses or events, interesting hiking tracks or the following of a character in the user's favorite book? Depending on these scenarios, different data sources have to be used and processed with a variety of techniques of geographic information retrieval to make sense of potential location data in unstructured documents.

In this chapter, local search is always used in the geographic sense. The search for personal files on a local machine, i.e., desktop search, is not discussed. The terms location-based or geospatial search applies to the retrieval of location references from documents. In this sense, Local Web Search means the search for information with a geographic aspect, not necessarily about the surroundings of a user. This aspect is more important in mobile search and is covered in Section 4. Local search is understood as an application of a specialized or vertical search engine. These complement or replace common general search engines for specific tasks (Lewandowski, 2009; Steele, 2001). Most cases of vertical search have in common that they implement a thematic restriction and employ topic-specific processing, indexing and search, taking the characteristics and potential metadata of its subject into account. Pure database systems are regularly not covered by the term. The application of general search engine principles to a domain-wide search or sources restricted to a set of domains usually does not meet the criteria for vertical search and is known as customized search. Some specific search engines such as aggregators use a similar approach but additionally exploit specific structures on domains to retrieve associated metadata or pages as structured content (Arasu & Garcia-Molina, 2003). However, for a general local search, the restriction to certain domains proves to be counterproductive. Local search is thus a vertical search as it concerns a nonobvious characteristic of data, requires special processing and indexing structures and allows interactions beyond the pure keyword paradigm.

The most interesting aspect of Local Search from a research view is that location is special in that it needs a different semantic than most other attributes and that is has a strong correlation to the real world, implying a massive usefulness for ist users. Another is that geography is not well mapped in purely textual search and needs adaptations throughout a search engine architecture (Ahlers & Boll, 2007). Geospatial Search therefore is a type of entity-oriented search in which the entities in question are not necessarily directly available in the source documents and have to be semantically extracted as described in Section 2.2.

Location now is a driving force behind many research and commercial search activities. Even while efforts to a more metadata-aware Web are underway, a wealth of geospatial information can already be extracted from today's Web and enables Local Search that is being used regularly to search, explore, and visualize geospatial Web content according to its location semantics. The prerequisite for this is to understand and be able to exploit the geospatial character of the Web to fully utilize this dimension of search. Figure 1 gives an overview into four processes identified in (Boll & Ahlers, 2008) as challenges for extraction and utilization of spatial data on the Web. The processes and subprocesses chiefly concern discovery of resources; understanding of documents and their location semantics as well as data mining; augmentation and enrichment of resources by correlation and integration with other sources; and finally exploration and search by ways of visualization and user interaction. The process cycle also includes a loop back to the beginning, similar to Vannevar Bush's vision of the Memex (Bush, 1945) that would allow users not only to access data but also to record and reuse their own "trails" through the information space that could then be published to strengthen or create connections in the data and create or correct information. This cycle also forms a frame for most activities in the context of Information Retrieval (IR) and Local Search.

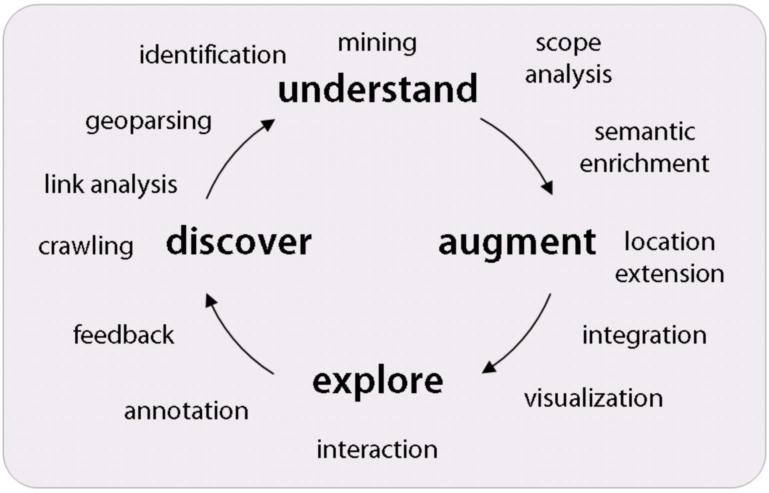

Fig. 1. Processes for interacting with data in a Geospatial Web (cf. (Boll & Ahlers, 2008))

2. History and Overview

Web search engines today are something taken for granted by most of the online population. With the enormous growth of the Web, it became clear very early that manually curated directories could not keep up with the growth and only automated systems could gather and open up the vast content of the Web.

Today, full-Web search engines are an impressive feat of software engineering, incorporating technologies from a wide variety of fields of computer science (Olston & Najork, 2010; Croft, Metzler, & Strohman, 2009; Levene, 2006; Arasu, Cho, Garcia-Molina, Paepcke, & Raghavan, 2001), with an emphasis on Information Retrieval (Rijsbergen, 1979). These techniques usually only support query types that were anticipated in the processing and indexing. For example, if textual document content is indexed, keyword queries are possible. However, this does not allow for geospatial distance queries, which would require understanding and indexing additional characteristics of the document as well as an adapted query engine and result presentation, basically mandating adaptations in all parts of a search engine (Ahlers & Boll, 2007).

2.1. Local Search

Efforts towards Local Search have a long history. As an intermediary step towards local Web search, location-based services (LBS) describe systems that maintain a database of mappable items, often as redacted databases of points of interest (POIs) or more general services. The roots of LBS and geospatial search lie in Geographic Information Systems (GIS) that manage and process geographic data and physical features of the world, mostly from a geographer's or cartographer's perspective (Heywood, Cornelius, & Carver, 2006; O'Sullivan & Unwin, 2002). While GIS can represent a wide variety of feature types such as points, polygons lines, or area shapes, most LBS only provide databases of POIs, with an emphasis on the point characteristic. Although LBS thus are less powerful in expression, they allow for a different class of application and often use GIS systems or technology in the background.

Making the step from GIS over to GIR (Geographic Information Retrieval), in a seminal paper titled "Spatial Search and Browsing", (Larson, 1996) discusses how to access geospatial information hidden in digital libraries and enable its search for this new class of retrieval. Larson is mainly concerned with the major issues of searching and browsing geospatial information, i.e., the user side of Information Retrieval. He presents different types of search still rooted deeply in GIS, such as ranges, polygons etc. and presents various search and browsing interfaces. However, different from Larson's work based in digital libraries, the Web is no curated library where all relevant metadata is already annotated. More details about this field initiated by Larson is given in the next section. Making use of the special characteristics of location, the user interface for geospatial search can provide a richer experience than common link lists by pinpointing results on a map so their spatial distribution and location can easily be assessed by the user.

Local Search in the Web context started hitting of in the first decade of the 21st century. The distinction between Local Search and the more specific Local Web Search that takes its data from the Web will examined more deeply in Section 3. Among one of the first to bring traditional yellow page book online, Yahoo! Initially debuted their yellow pages in 1997. Today, most yellow page providers also have a Web presence to give access to their data. For most of local search, the yellow page metaphor is still used in many places while on the other hand, the Web itself provides even more data previously only available in directories databases (Himmelstein, 2005). As one of the first services using a map interface, Google Local[1] was introduced in 2004 and provided local business listings and a map featuring directions and routing. The service has since been merged into Google Maps that was introduced in 2005 with the then-new service of satellite imagery. At the same time Google Earth was introduced, bringing the map interactivity to the globe. This also spawned a variety of new user-generated content, as users were now able to provide their own geospatial datasets in the KML language to be shown on the globe or would start creating mashups of geospatial data and available map APIs on the Web. With the growing perception of geographical metadata and coordinates, many new startups were growing rapidly and also users found its use due to the new availability of analysis and visualization services, so that location metadata was added in many places, such as images, blogs, wikipedia, etc., leading to what is known as the Geospatial Web (Scharl & Tochterman, 2007). In one notable example, flickr images were made searchable by location (mappr.com, since defunct) and after Yahoo acquired flickr, it provided many new applications for geotagged images, such as toponym tag disambiguation and placing (Serdyukov, Murdock, & Zwol, 2009) or place and event detection and geospatial clustering (Rattenbury, Good, & Naaman, 2007). Similarly, most location-based search systems are now also available in a mobile version due to the powerful mobility wave (cf. Section 4).

While in the beginning, local search mostly just took data from yellow page providers, in recent years this has changed to include a massive amount of Web resources as well. Naturally, this makes it much more of a challenge to provide accurate and high-quality data, which is usually assured by quality control processes that employ multiple heuristics and thresholds to filter and aggregate the raw Web data. Due to the increase in location data and the increased desire by business owners to manage their local search presence, mechanisms were added by search providers to allow them to directly manipulate entries, such as Place Pages by Google that was launched in 2009[2].

---

1 http://www.google.com/about/corporate/company/history.html

Around that time, local search results were also starting to be integrated into the organic Web search results in the form of universal search. Most of the local search providers also allow for some version of paid content or advertising on their pages and in the result sets by search engine marketing, a market expected to grow massively in the coming years[3]. Taking this into a slightly different direction, foursquare in 2010 published a system of location-targeted offers that users could participate in by publishing their visit to a place. Finally, social networks and Web 2.0 services are also starting to embrace location to either attach it to entries or allow users to share and comment on their locations.

Location-based search today ranges from simple POI databases over yellow pages, specialized search engines, mashups and user-generated content up to Web-enabled GIS systems and full Web search. There is a tradeoff between curated databases and Web search. The former can provide very accurate, manually edited information with high credibility. The latter can provide a high information depth, a much wider variety and current information at lower cost and much shorter renewal cycles, but they have to be extracted with sophisticated technologies and may not always be of high quality. In conclusion, both systems have their benefits and their applicability depends on the application's requirements. In recent years, many hybrid variants emerged that operate both on curated databases and the rich resources of the Web.

2.2. Geographic Information Retrieval

Geographic Information Retrieval is a specialization of Information Retrieval (IR) that aims to add the semantic dimension of location and describes the identification, augmentation, and processing of geospatial references from documents and the means to provide access to them (Purves et al., 2007). The challenge of GIR lies in the improvement of search by "treating spatial special". As the most prominent data source, most research is concentrating on Web data (Buyukkokten, Cho, Garcia-Molina, Gravano, & Shivakumar, 1999). One major aspect is toponym detection, the extraction, disambiguation and verification of potential location references in documents. (Gräf, Henrich, Lüdecke, & Schlieder, 2006) and (Leidner, 2008) provide good overviews over toponym extraction. A query for "castles in Germany" should not only find pages that contain the keywords, but also find castles in individual cities within Germany, a result that only becomes possible when a geographical understanding of the query is attempted. Following, the found terms have to be grounded to an actual location in the toponym resolution, usually by the use of geographical thesauri, so-called gazetteers (Hill, 2000; Souza, Davis, Borges, Delboni, & Laender, 2005).

Taking the step towards database-driven systems, many sites on the Web give out location information about entities in a structured way. By using dynamic template-reconstruction systems or similar approaches, this information can then be restructured again (cf. (Arasu & Garcia-Molina, 2003; Cafarella, Madhavan, & Halevy, 2008)), which of course can also include location references. In many cases, the location references are physical addresses. The extraction of these structures is discussed in (Ahlers & Boll, 2008) Apart from the commercial local search engines, which do not reveal much about their inner workings, research prototypes of geospatial Web search engines have been described in the literature and offer a good overview, e.g., (Lüdecke, 2010; Ahlers & Boll, 2007; Markowetz, Chen, Suel, Long, & Seeger, 2005; Markowetz, Brinkhoff, & Seeger, 2004).

2.3. Search Scenarios and Information Needs

The information needs of users of local search naturally differ from normal Web search. For the latter, (Broder, 2002) builds up a taxonomy of user intentions covering informational, navigational, and

2 http://www.google.com/places/

3 http://www.kelseygroup.com/press/pr090224.asp

transactional. A similar pattern can also be found for local queries (“I want to know about a place”, “I want to go to the place”, “I want to do something at a place”). (Battelle, 2005) among others describes two types of local search, recovery and discovery search. These are the search for the location of a known place (or the name of a known location) and the search to find an unknown place with certain characteristics, respectively. Another use case is exploratory search (Marchionini, 2006) where a user has no specific result in mind but just wants to explore the surroundings. Most often, discovery search is implicitly assumed. This is because even while the intent and the query are different in recovery search, search engines developed towards discovery also allow for recovery. In the case of local search this is a bit different. Recovery and exploratory search are often hindered, since there is no mode that would put all results onto a map. So to find the name of an entity at a known location most often a search term has to be given. Similarly, a general overview is often only enabled by an additional map layer showing few, high-visibility places. In most cases, a query can be deconstructed into a conceptual and spatial aspect. Hotels in a city, restaurants near the coast, the next gas station, the three nearest airports, historical information about a part of town, the spots of nightlife in a city are examples for this query type. The spatial aspect is most often expressed by a constraint to a certain area or a distance from a center by means of, e.g., a circle filter (R. Jones, Zhang, Rey, Jhala, & Stipp, 2008). More spatial queries are exemplarily described by (Larson, 1996), who names, e.g., polygon boundaries, rectangles, buffer zones, or first-k. Exploring the aspect of mobile users, (Mountain & Macfarlane, 2007) describes spatial filters such as spatial and temporal proximity or prediction of future positions. This takes into account that users often have a notion of distance of search results to their current position. This could mean a diminishing importance by growing distance (order by nearest results). On the other hand, constraints to a queried region are expressed (hotels in Berlin, which either equally selects everything within Berlin or orders it from the city center out). Additional ranking can take importance of a result into account (Ehlen, Zajac, & Rao, 2009). Depending on the situation of a user, more query modifiers apart from mere distance can play a role (Ahlers & Boll, 2009). For example, in a car scenario, results in the direction of movement should be preferred, for example in a search for gas stations or restaurants. A further aspect is the relevance of results to a user query and the subsequent ranking (Markowetz et al., 2005; Purves et al., 2007). More specific aspects are the spatial spread and selection of results depending on their number and distribution (De Sabbata & Reichenbacher, 2010) and spatial diversity (Tang & Sanderson, 2010) which aims to provide results spread over multiple locations if the ranking would otherwise only produce few clusters. Other studies have looked into the intention behind local queries and shown that these cover a wide area of topics (Amin, Townsend, Ossenbruggen, & Hardman, 2009; Gan, Attenberg, Markowetz, & Suel, 2008; Kamvar & Baluja, 2006). These topics are mostly similar to those in the Web interfaces, but exhibit a different distribution skewed towards local results. (Henrich & Lüdecke, 2007) is recommended for a further overview of spatial information needs.

3. Current Local Search Engines

After the preparatory technical background in the previous sections, the article will now discuss current local search engines and compare the used technologies. It will not give a full formal comparison, but use the exemplarily selected search engines to highlight the technologies behind the scenes, how they are used and combined and how they influence and enable the local search experience available today. The subjects will mainly be systems for the US and Germany, owing to this researcher‘s experience and the good data availability to discuss state-of-the-art systems.

Unfortunately, many statements regarding algorithmic details and actual use of technology have to remain vague, as there is little published information about the actual workings of the backend system. Most publicly available information only deals with visible features in the final product, but this sometimes allows to draw inferences about underlying data collections or algorithms[45]. However, only little current research is available from the mayor players on local search, with some noteworthy examples given throughout this article, such as (M. T. Jones, McClendon, Charaniya, & Ashbridge, 2007; Xiao, Wang, Xie, & Luo, 2008; Kamvar & Baluja, 2006). For Web search, the companies are very active in publication, showing that search still is a problem far from being solved. Due to these restrictions, the discussion will present possible data sources and processing as well as use interaction and search modalities based on available sources and inferences of use technology.

In all examined search engines, a map view is used, either as the main interface or at least as a viewer for the locational aspect of the presented data. This is unsurprising, as maps are the most intuitive presentation of geospatial data (Downs & Stea, 1977; Larson, 1996; Lynch, 1960). In most cases, the maps and the satellite imagery is not produced by the search engines themselves, but bought from specialized providers, and then preprocessed for the use in local search. While satellite imagery is stitched and color corrected and undergoes other image manipulation steps, the vectorized map data is cleared and then rendered in a way deemed most applicable by the search engine provider. However, there are of course many smaller services apart from he big ones that simply use those maps via API calls. An area in which the search engines have been active themselves is the addition of eye-level street imagery such as Google Streetview[6] (Anguelov et al., 2010) and Bing Streetside[7], in the case of Bing with added 3D reconstruction to match arbitrary photographs of the street scene into the view (Snavely, Seitz, & Szeliski, 2006).

What may be slightly more surprising is that the map is often already centered on the user's location or a location keyword for the search is recommended. This feature is an aspect of content localization in which providers try to guess the user's location to adapt the content or the interface. The lookup is done by the IP infrastructure of the Internet. The user's IP address is looked up in geo-IP databases which show which organization owns an IP range. The lookup on simply a country level is simple, as IP ranges are assigned on a per-country basis. For a better accuracy, other information has to be used. For example, universities can often be located at a rather exact location as they have a limited IP range and the administration data shows their address. On the other hand, many carriers have a huge block of IP addresses for their home users, but have them divided internally by region so that often, a position down to a city level is possible. This is done by a combination of automatic and partly manual processes to maintain the databases (Freedman, Vutukuru, Feamster, & Balakrishnan, 2005). It can be tried out with many services[8] and also easily integrated into applications with available APIs. This location happens solely on the infrastructure level and does not need any user interaction or special software. To provide more accurate location to applications, geolocation has recently been integrated into browsers that can then locally use WiFi fingerprinting or GPS and also allows users to manually give their location to a service[9].

---

4 Google Lat Long Blog http://google-latlong.blogspot.com/

5 Yahoo! Geo API documentation http://developer.yahoo.com/geo/

6 http://maps.google.com/help/maps/streetview/

7 http://www.microsoft.com/maps/streetside.aspx

8 e.g., http://www.geoiptool.com/, http://www.hostip.info/, http://www.tracemyip.org/

9 Geolocation API Specification, 2010, W3C http://www.w3.org/TR/geolocation-API/

Another aspect of local search engines is their usability, an aspect that is left out of this article, but was widely studied, e.g., in (Lewandowski, Nesbach, & Mikley, 2010), using the example of Google, Bing, and T-Online for Germany. Other work performs user studies (Amin et al., 2009) or examines search engine logs for locational preferences (Kamvar & Baluja, 2006; Henrich & Lüdecke, 2007).

3.1. Yellow Pages

This section discusses yellow page searches by using the two examples of yellowpages. com[10] for the US and Gelbe Seiten[11] for Germany, as these are among the most popular for the respective countries. Yellow pages traditionally are bound directories of business' phone numbers and addresses. The step to publish such database-like information on the Web is small and most publishers have done it. However, there are wide differences between different services, with some only providing basic search for business name and city; and others building powerful portals around the data with hierarchies, tagging, connections to other databases etc.

The raw basic business details are usually collected by the publishers directly while selling the space in the directory. Additional information can be added. Often it remains unclear whether this was already collected from the entity or later merged in from other databases. One of the features of directory-style data is that it is associated with a line of business that is maintained as a controlled vocabulary by the publisher. This vocabulary is usually ordered as a hierarchy in the Web interfaces. However, sometimes the hierarchy only exists for the classification names, but as a look into a paper yellow pages shows, the category names are not really ordered hierarchically. The classification itself is done by the yellow page publishers and depending on the workflows used for the Web search, the categories provide a real selection or are only used for a keyword search. In the latter case, their selectivity is reduced. In the examples, Gelbe Seiten uses only one line of business while yellowpages.com uses categories more like tags.

Both examined services provide a hierarchy of classification terms, either by autocomplete of the search string or as a separate pane besides the map view to get an overview of available entities in a region. A search is initiated in both by using a simple concept@location query that requires two different input boxes for the search keywords and the location, respectively. For the location input suggestions are given via autocomplete. These draw from a gazetteer, a geographical thesaurus, that models knowledge about place names and their relations (Hill, 2000). In some cases, the used data sources are not in sync. In one service, the location of the author's university in Tegucigalpa, Honduras was already filled in by user geolocation, but the service could not provide any results in that area. In Gelbe Seiten, if no match could be found, it automatically increases the search radius to incorporate other nearby places, while yellowpages.com only shows an empty page with a minimal hint, but no further assistance. As an obvious evidence of merged data sources, yellowpages.com treats hotel queries differently and uses listings and additional information from hotels.com to answer such queries.

The basis of the services remains a business directory, even if that information is sometimes augmented with other sources. Usually, yellow pages do not find information that is not in a phone book, so no points of interest or sights can be found if they have no business interest behind them. Examples for such non-business POIs are sights, neighborhoods, famous streets or harbors, beaches, lakes, abandoned ruins etc. but also phone boxes, ATMs and similar. For example, YP does not recognize Times Square in itself, but rather gives out business listings that have the street in their

10 YP.com http://www.yellowpages.com/

11 Gelbe Seiten http://www.gelbeseiten.de/

name. Similarly, Gelbe Seiten does not recognize well-known places for Germany, such as the Brandenburger Tor in Berlin, but has added categories for entities such as parking lots or police stations. In conclusion, yellow pages provide no real Web search but rather a search interface on the Web to their own edited (and paid-for) high-quality data. Yet, with their data they provide a basis for many other services that use them to bootstrap location-based services[12].

3.2. Map-based Local Web Search

It has long been discussed that eventually the Web will contain all relevant location information and make yellow pages obsolete (Himmelstein, 2005). Yet, even if indeed a majority of places can be found on the Web, there are still many places that cannot be found this way (studies estimate between 5% and 25% of businesses in the developed world have no Web presence, with this number being much higher in developing countries, e.g., (EUROSTAT Press Office, 2008)). Furthermore, as opposed to manually redacted yellow pages, the identification of reliable Web sources or the derivation of credibility indicators over multiple pages as well as the reliable extraction of data can be a difficult process (Fogg et al., 2001).

The first attempts at generating local Web search were aimed entirely at the Web and the extraction of address patterns. After this was found to be not feasible, a second generation used licensed yellow page data almost exclusively. The current forms found today are hybrids in that they are based on a combination of different data sources. Data is usually taken from yellow pages directories for each country where the service is offered; freely crawled Web pages from the general Web index that are searched for verifiable addresses (Ahlers & Boll, 2008); extracted structured data from certain Web sources (Cafarella et al., 2008; Chang, Kayed, Girgis, & Shaalan, 2006); submitted data of business owners that claimed an existing listing or created one; and finally user-generated content and crowdsourcing of updates and corrections[13]. The individual search engines of course have their own approach to the source selection and merging, but this describes the general sourcing process[14]. Some sources are treated special in that they have a higher perceived credibility and/or allow a more structured extraction of information. This is mostly the source of reviews, ratings, prices etc. on hotels, restaurants and similar results.

The local Web search is basically an entity search. It abstracts from Web pages and rather displays results based on the entity that is described in those pages. To achieve this, the entities have to be merged after all their source documents are found in the Web index via entity extraction. All pages in the index that could contribute to an entity are scanned for properties that can be added into the entity or that can serve to corroborate already present evidence on an entity's properties. Merging entities can be supported by external domain knowledge in the form of gazetteers, yellow page entities, product name databases, phone area codes (Amitay, Har'El, Sivan, & Soffer, 2004) etc. (Lee, Kitayama, & Sumiya, 2008). To define which entities and properties are merged, similarity measures and thresholds are defined on addresses, for non-exact entity name matching and other properties. Additionally, frequency analysis of sources and PageRank-like functions can be employed over the data. The signals which inform the ranking are not published by the search engines, yet they can be guessed or derived

12 Directory data is not only printed and disseminated by their own Web site, but can also be licensed by other businesses. Major search engines use this data as a tool to ground some of their found local entities.

13 This can be done easily in certain map views. One step further is taken by Google Map Maker where the underlying map can be changed as well, similar to OpenStreetMap.

14 cf. http://maps.google.com/support/bin/answer.py?hl=en&answer=7103

partially[15]. For the final display of results, object-level ranking methods (M. T. Jones et al., 2007) are added to the rest of the signal mix for the overall rank. This shows that the process is much less yellow-page-centric than in recent years. It can also be attributed to the fact that many businesses cut back on presence in classic yellow pages and rather try to be found on the Web and in many Web databases. This means that entities that are in the yellow pages are part of local search, but their entry grows in information depth if there are other sources adding to their entry and describing them in more detail. Also, entities that are described only in the Web can be part of local search results if they are present in sources with higher credibility. An entity that is only described in its own, poorly linked Web page without an entry in yellow pages, Web directories or social networks will probably not show up in local search at all. If it does, its ranking will be rather low and it only show up in later result pages.

The examined local Web search engines are Google Maps, Yahoo! Maps, and Bing Maps as these are the most commonly used[16] and are integrated into the largest general Web search engines[17]. Starting with the query interface, all allow to pose a textual query, with Yahoo and Bing using two separate fields for a concept@location query, while Google uses a single input box. For the latter case, the separate inputs for what and where are automatically extracted from the query by a geocoder/geoparser (Goldberg, 2008) system adapted to queries (Hines & Abou-Assaleh, 2009), making input much easier for the users. This is especially interesting in the context of switching between different search types such as Web search, local search and back. This is offered by respective links in the interface, but is not fully supported by all systems. While Google allows seamless transitions, Bing only takes over the keyword part when switching from a local to a Web search, but can separate keywords and locations when switching back. However, this works only when the location part comes last. Yahoo keeps the location part in the keyword and additionally assigns the location when switching to local search, leading to no results, but showing a geospatial capability as well in identifying the location part; while the other direction correctly assembles a search string.

The result page looks similar in all services and consists of a map with a result list. The map zooms automatically to a level depending on the spatial distribution of results. The map background is diverse, either because different map providers are used or because the map rendering is adapted. The left-side view gives a paginated list of results. These – usually 10 – results are shown on the map with pin markers. Users have to page through them to see more results. The results at least carry name and location and may show additional structured information such as phone numbers, distance to the map center, homepage, ratings, or reviews.

For some services, the query can be refined by moving or zooming the map to the area of interest. Bing actually removes search results from the map in some cases and cannot find results when the location part was left empty, even though research exists by the company on how to analyze user queries and panning movements on a search result map (Xiao et al., 2008). Google Maps instead allows giving a query without a location part. In that case, but also for other cases where a user wants to drill down into a part of a region or city, the location selection is given by the current view of the map, which also allows dynamic selection and re-ranking of results based on the chosen view. Yahoo

15 cf. David Mihm, Local Search Ranking Factors, 2011 http://www.davidmihm.com/local-search-ranking-factors.shtml

16 comScore Webinar, State of Local Business Search, 07-20-2011

17 comScore June 2011 U.S. Search Engine Rankings, http://www.comscore.com/Press_Events/Press_Releases/2011/7/comScore_Releases_June_2011_U.S._Search_Engine_Rankings

lies in the middle, being able to give results based on the map, but only shows results clustered around the center of the map. Especially for zoomed-out maps, this hides much information. However, for more close-up zoom levels, the search radius is not only controlled by the current rectangular map view, it is also possible to change the radius of the circular search environment (Figure 4).

No aggregation of results is used in the map views, such as in flickr maps that shows for large views only an aggregated view of the numbers of pictures and distinguishes them in higher zoom levels. However, to not be dependent only on paginated result lists, Google chose another way of displaying overall result density. Results other than the ones currently shows in the paginated view are shown as small red dots. This has the twofold effect that they are less prominent on the map, but also allow to show much more information without overlap. The result selection and ranking changes dynamically with pan and zoom operations on both the major and the minor markers (Figure 3).

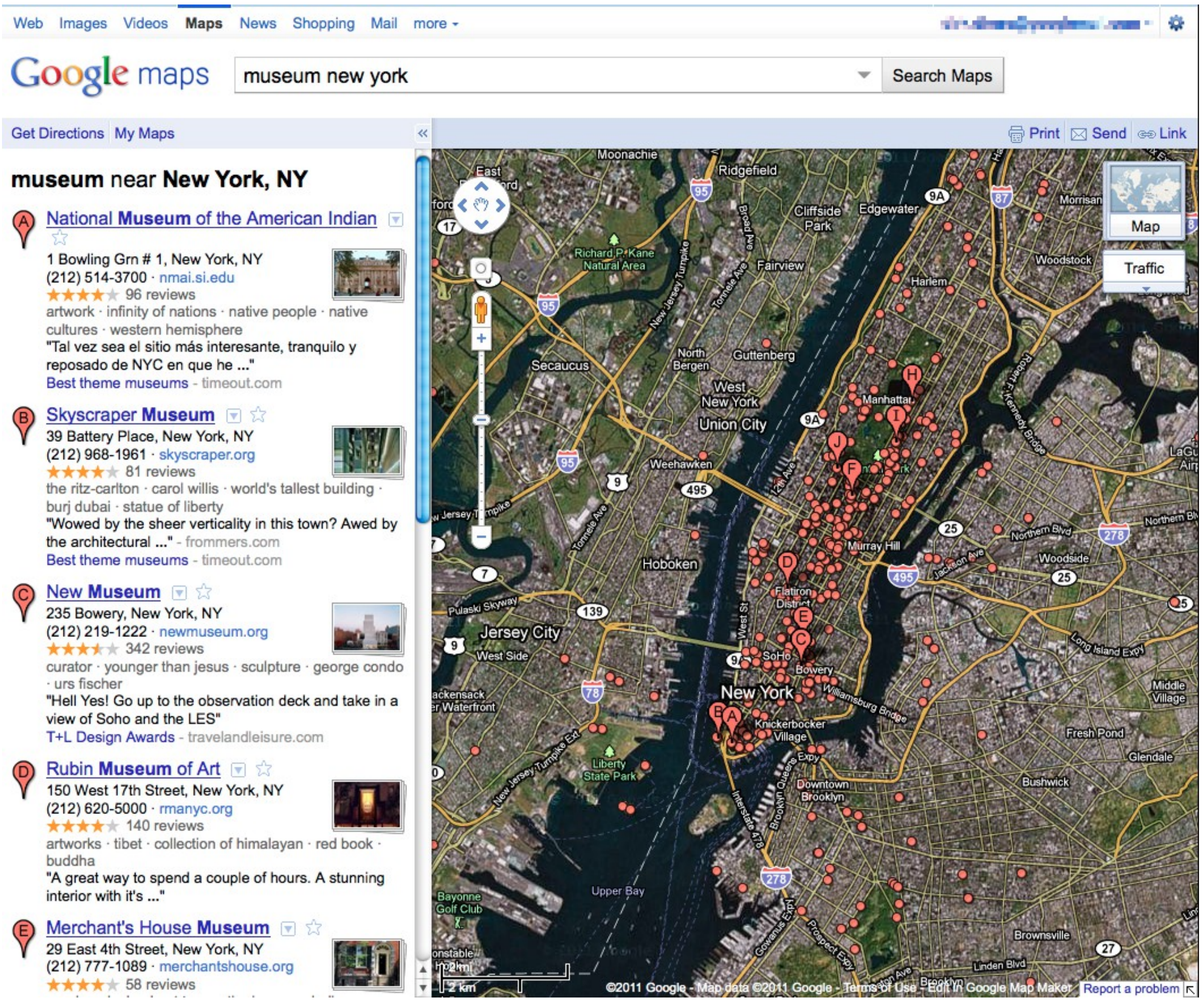

Fig. 3. Google Maps results of a search for museums in New York

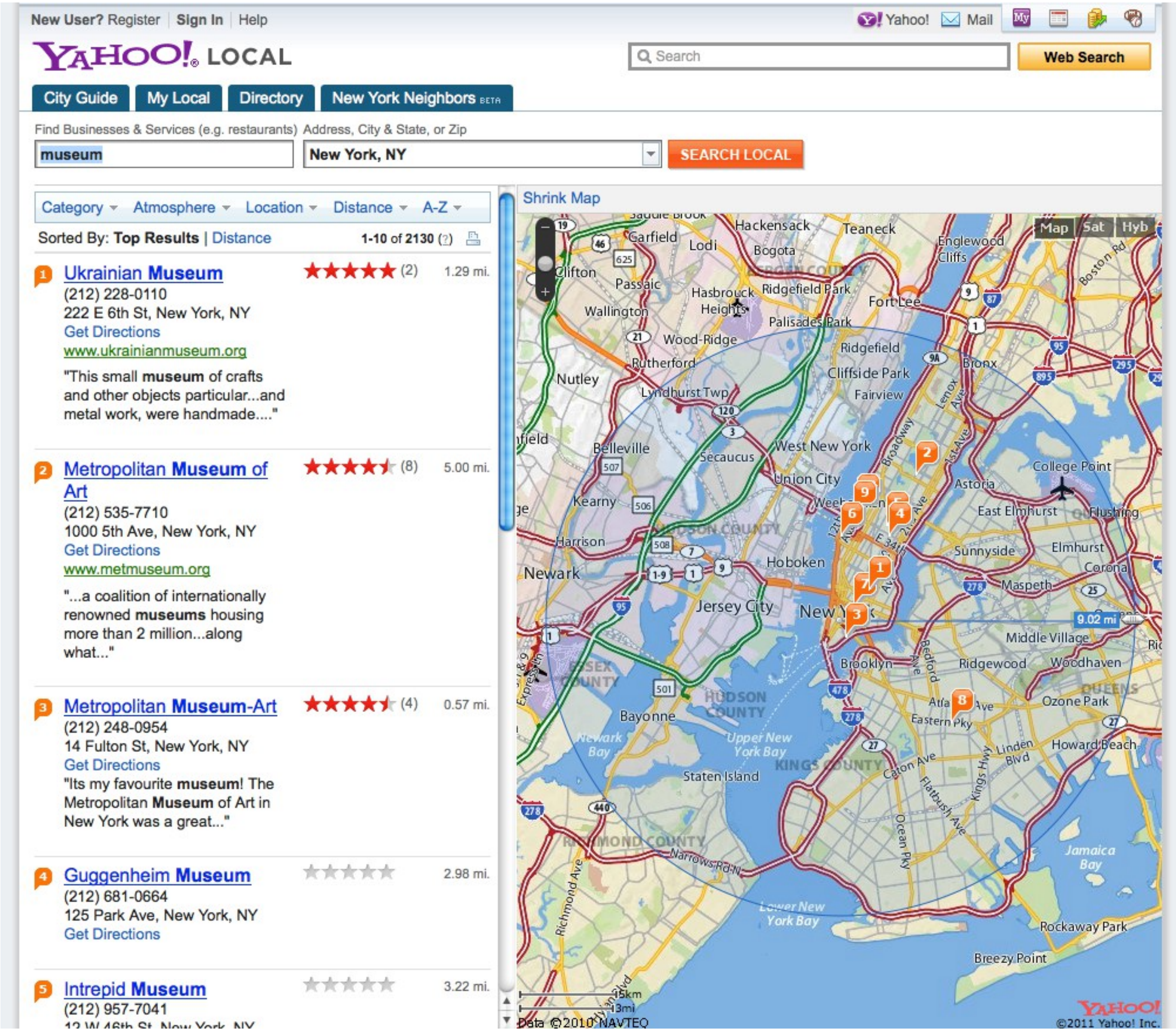

Fig. 4. Yahoo! Maps results of a search for museums in New York

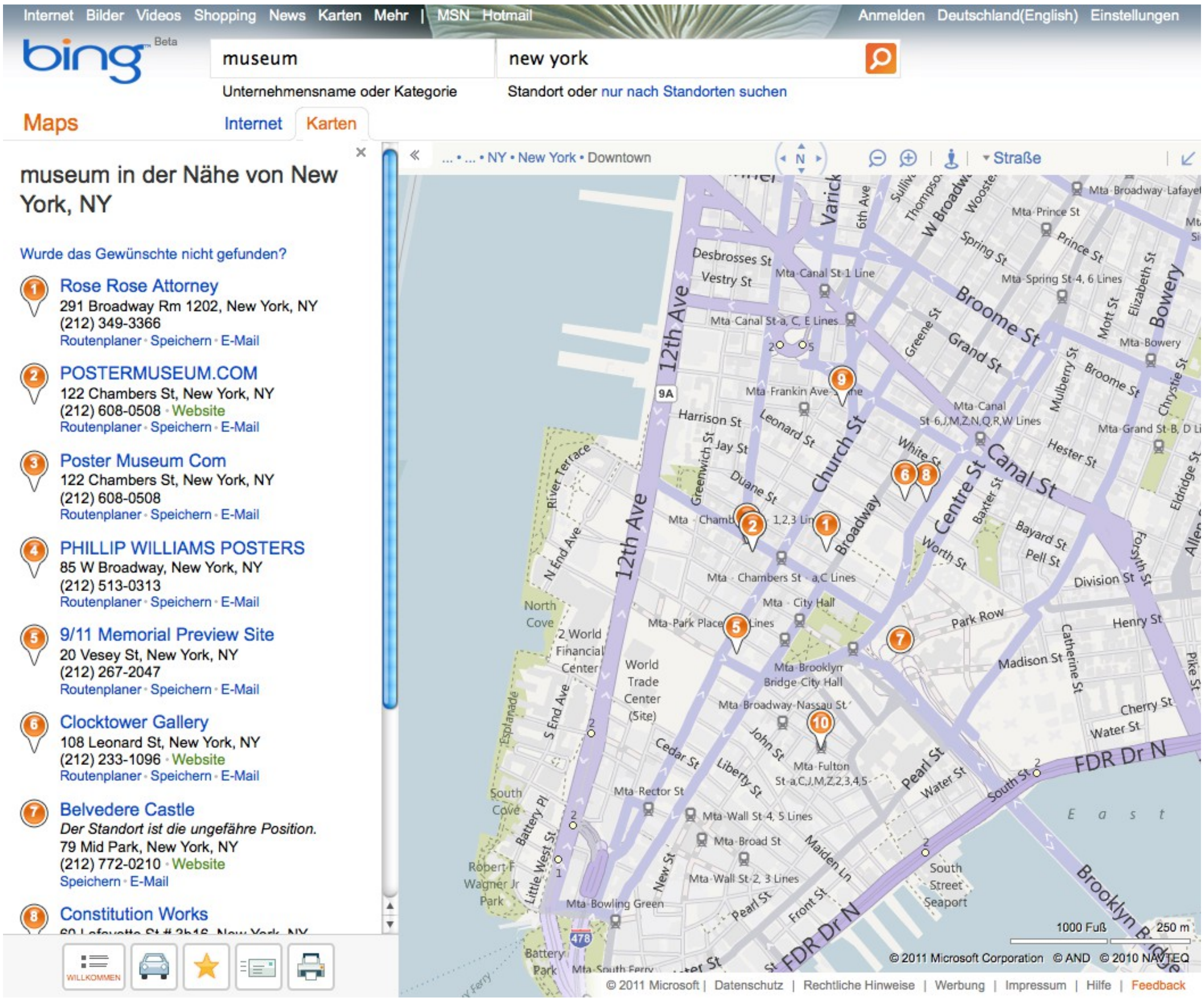

Fig. 5. Bing Maps results of a search for museums in New York

The detail view on a single result shows the difference between the source selection and integration of the search engines. Bing (Figure 8) provides the most basic entity information with the name, ist address, phone number and homepage and its location on a map. It also gives the neighborhood, which can be automatically taken from a gazetteer, categories (it is not clear where these are taken from, either the homepage or a business listing) and provides route calculations to this point. Yahoo (Figure 7) presents an entity as an overlay window on the map so that other results are still visible. The entity has the same properties as with Bing. Additionally, it contains reviews from its registered users as well as reviews crawled from the Web and a computed score, gives a location based on street corners, contact information, an abstract, and opening hours. Reviews and images are on separate tabs. The latter seem to by mostly crowdsourced from its own users. Google (Figure 6) shows the most information for an entity and visibly merges multiple sources for it. Details such as classification, images, opening hours, prices, parking, public transit information, upcoming events and more are displayed with an entity if that information is available, either from the Web or external sources. An abstract may be taken from wikipedia. Additionally, it provides a term analysis of reviews, extracting the most important terms (products, services, etc. such as “exhibition”, “collection”, ...) available and what users are saying about them. The images come from panoramio.com, a Google subsidiary and are selected as one representative photo and a list of further images. In a detailed source overview, all found sources, both reviews and common Web pages, are accessible in ranked lists similar to a Web search result, with the reviews annotated with a score and the source. Additional reviews from registered users are also presented. Above the map, it is also possible to navigate to the other results from the initial search. Users can initiate a search for nearby or similar places or save a place to their own maps. As additional service, it allows business owners to “claim” their listing to edit their own information. This is a rather huge amount of information, but it is arranged by sound information design principles from top to bottom with an order of importance to a user, thus remains usable.

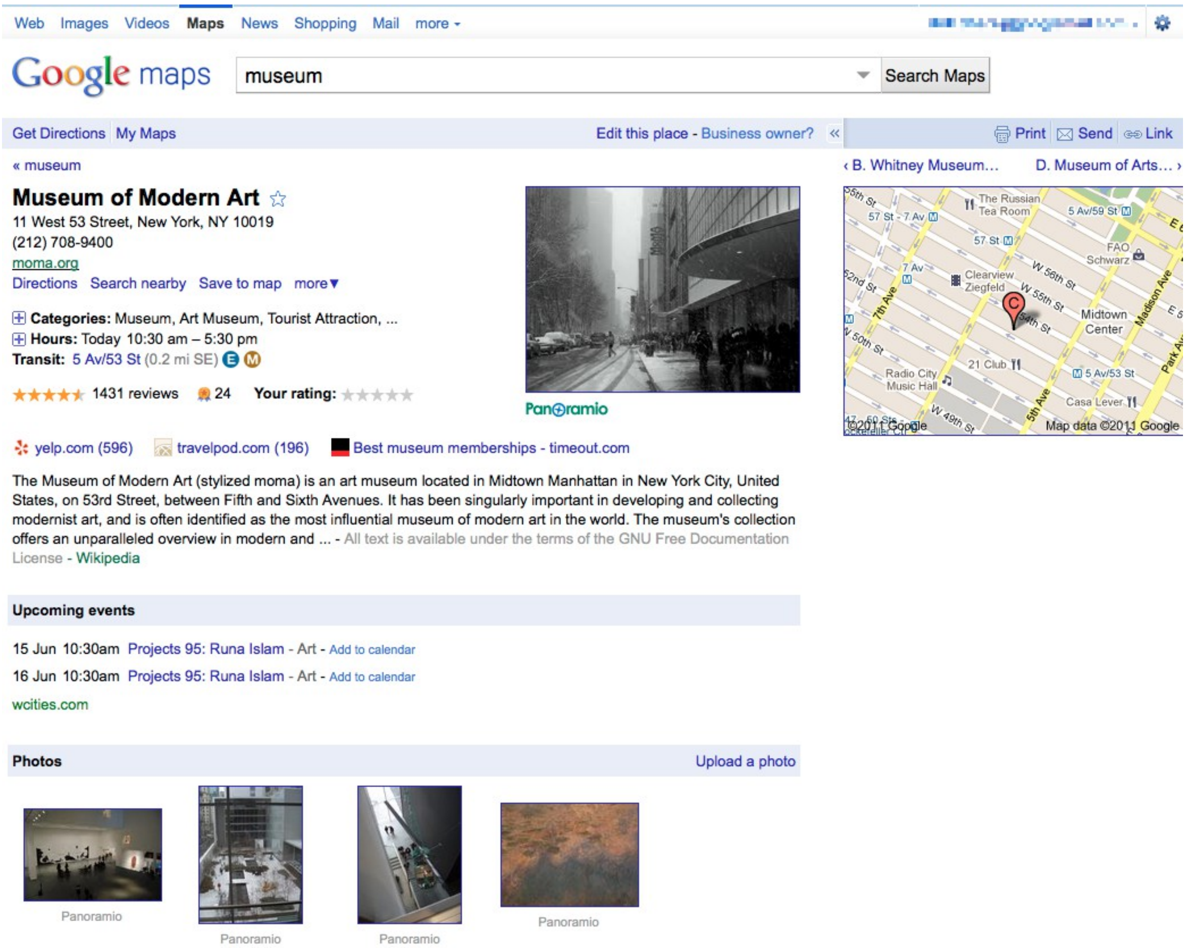

Fig. 6. Google Maps result detail for the MoMA, New York

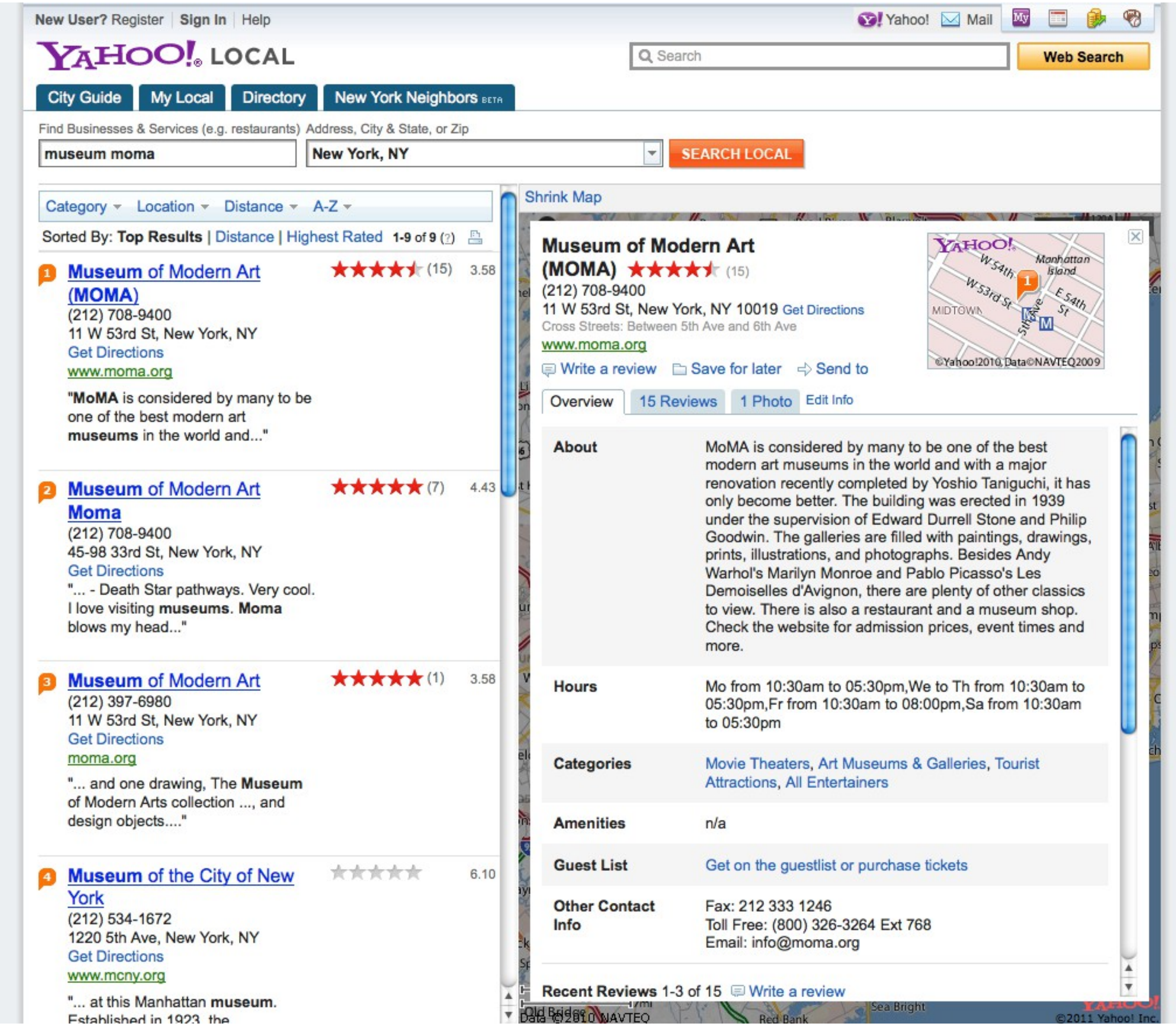

Fig. 7. Yahoo! Maps result detail for the MoMA in New York

Fig. 8. Bing Maps result detail for the MoMA in New York

For all examined systems, hotels are treated differently and often have more business information attached. This information is taken by collaboration from other specialized hotel databases or reservation systems. It ranges from a simple list of amenities in Bing over reviews, prices, payment information in Yahoo and up to listing of average prices and direct access to bookings by accessing multiple booking sites in Google. Additionally, when it detects a query for hotels, it unlocks additional query fields for a date range and can search for only hotels with available rooms on those dates. For other information that is not commonly available such as police stations, parking lots, telephone boxes or ATMs, external licensed databases are used and merged into the map view since this information is usually not crawlable on the Web.

Apart from just parsing pages for extractable addresses, names, phone numbers etc. by entity extraction, another angle has recently gained interest again. While common metadata in Web pages is currently mostly ignored by search engines due to search engine spam, structured annotations inside pages in the form of microformats, RDF, or similar approaches are starting to get used. Such structured data today is usually correlated with a reliability score to avoid spamming.

The way to receive more trustworthy data from users and page authors is shown by two exemplarily approaches. The first is Place Pages and the Local Business Center[18] that allows business owners to provide their own information, with the noteworthy example of opening hours that are manually added, and not extracted from freely crawled text. A second one is the use of Rich Snippets for local search[19], by allowing authors to add microformats to a page that can be read and used by the search engine.

The actual view and the information depth depends heavily on the available and used datasources. As such, there are many differences between different countries. In cases where structured databases are available, they are often integrated into the results, providing ore credibility and more attributes to entities, in other cases, only very basic information is available. A full examination of country-dependent results is out of the scope of this article, but would be interesting future work.

Geospatial search is not only used in the local search, but can also used in other speciality searches. For example, Google Books collects much metadata about articles and books, but also performs lexicographical analysis to find the most distinctive terms and phrases in a book. Using technologies long in the focus of GIR, it also perform toponym extraction to generate a map of places and toponyms found in a book as seen in Figure 9. How GIR is also used to improve general Web search is explored in the next section.

18 http://googleblog.blogspot.com/2009/09/place-pages-for-google-maps-there-are.html

19 http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=99170

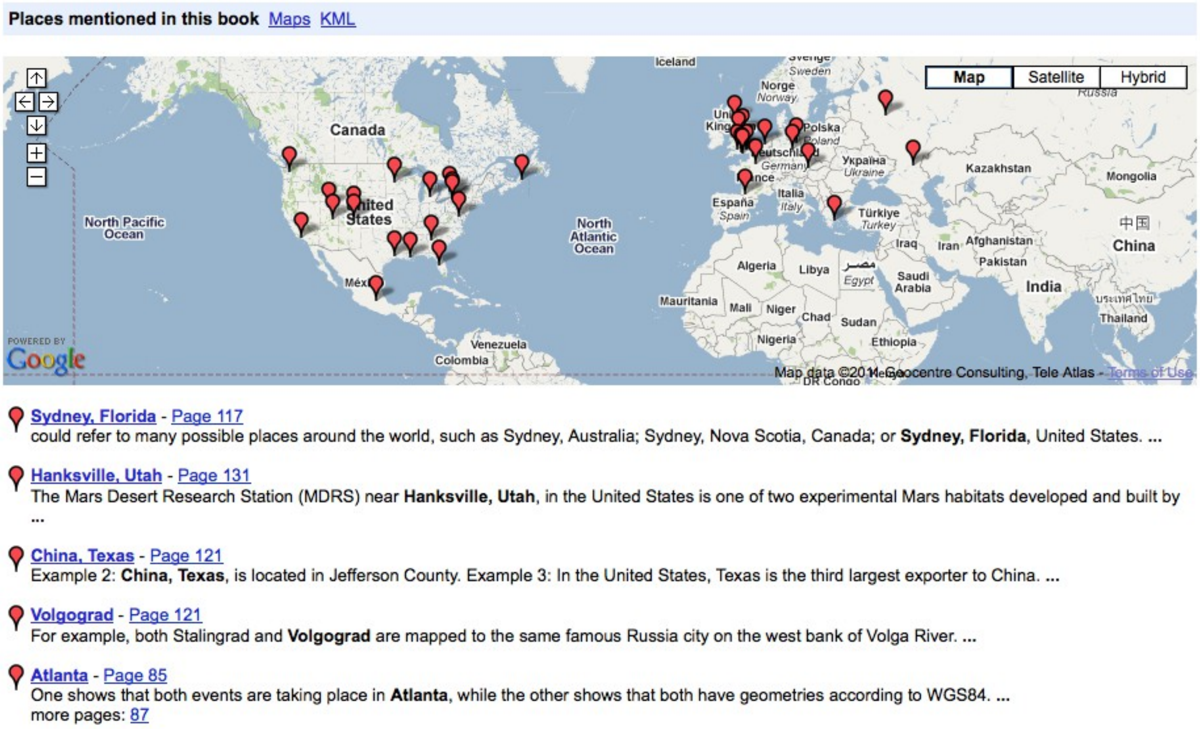

Fig. 9. Example of toponym extraction in Google books for the book (Scharl & Tochterman, 2007)

3.3. General Web Search

The reason that general Web search is included in these discussions is that in some instances it uses some geographic aspects in delivering results as well. Without any geospatial ability, general Web search can already deliver rather good results for certain local queries. Especially for non-business POIs or rather unique entities this works well. However, it can not deal well with ambiguities as little disambiguation of the query keywords can be shown to the user. So these queries need geospatial ability and possibly also a map view to distinguish between multiple places with the same name as discussed in the previous section.

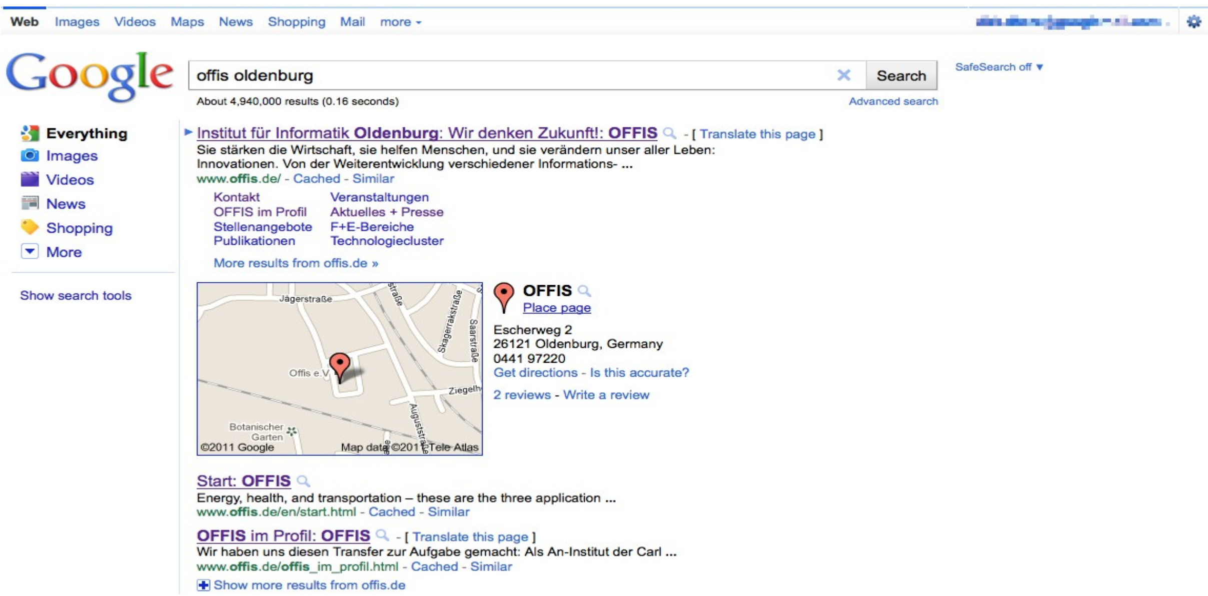

Fig. 10. Example of universal search results with a georeferenced entity

For the discussion we first examine Google Web search. The first, most obvious feature i the inclusion of some results from the maps search at the beginning of some result pages as shown in Figure 10. This so-called universal search merges results from the different areas (such as blogs, images, videos, maps) with the organic results of the Web search. The integration of local results is triggered most often with the combination of a keyword with a city name. The keyword needs to be local in a way, meaning that it generates a certain set of local results that are rather prominent and exhibit some clustering, e.g., the results have to be focused on one or two areas (for example, ”hotels in New York“ instead of ”hotels“). Local search result items that would be homogeneously distributed about the map do not reach the threshold needed to justify putting in a local result that needs to be considered as locally relevant. However, it can also provide location data for neighborhoods of non-business POIs. For the example of Times Square, NY, USA, it resolves the name by its gazetteer to the full designation. Also, it not only provides a map, but also georeferenced images from panoramio.com depicting the location. Additionally, some more specific location-oriented query offers are given (Figure 11).

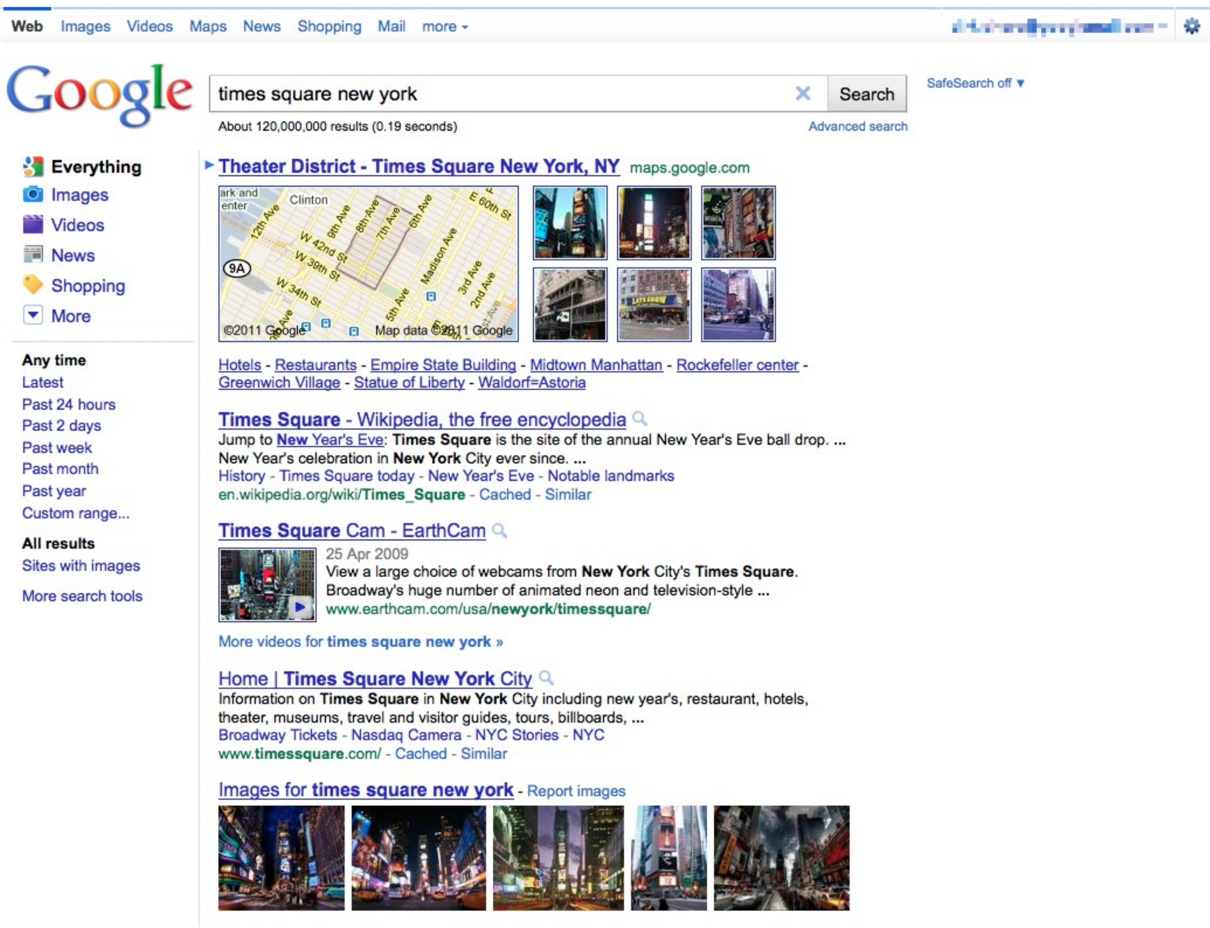

Fig. 11. Example of universal search results with a georeferenced sight and georeferenced images

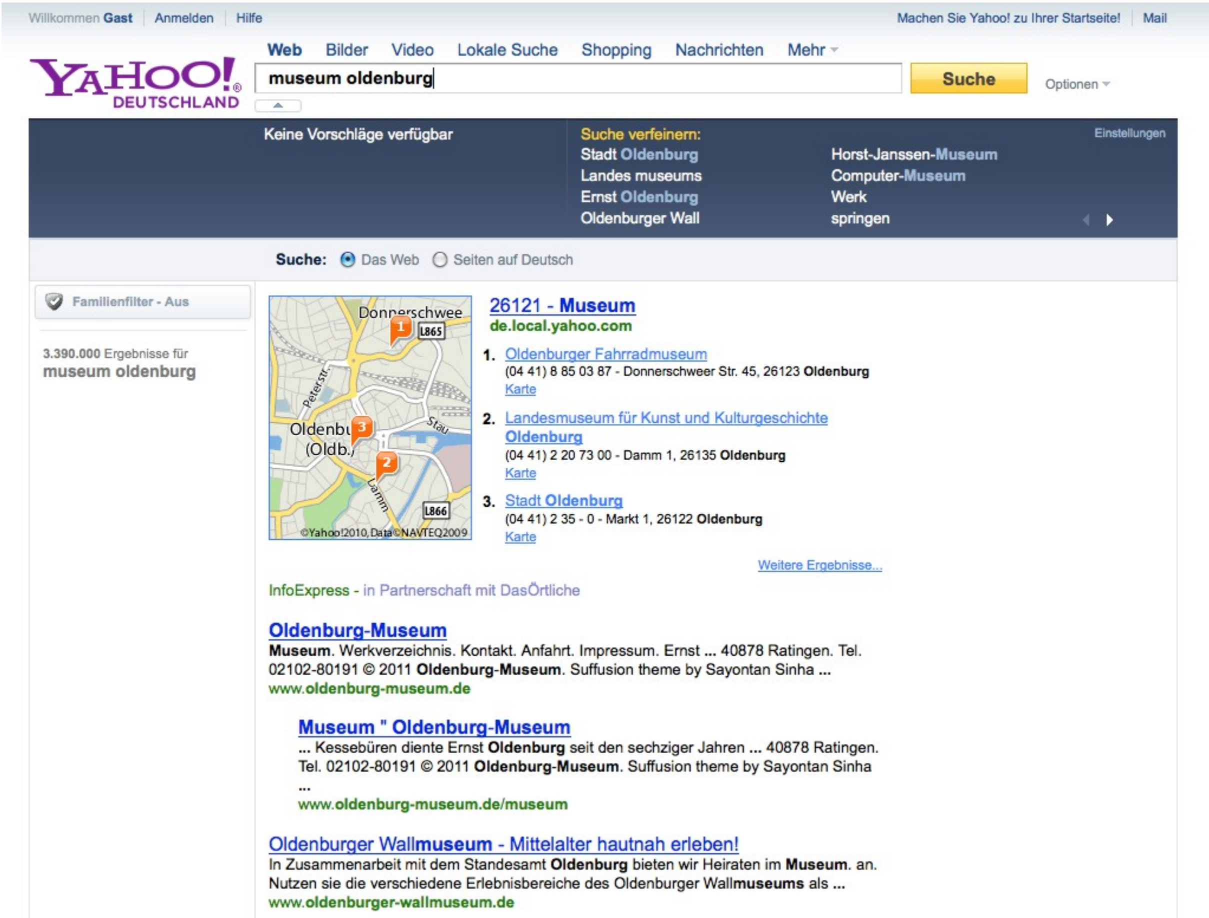

Fig. 12. Example of Yahoo! universal search with georeferenced results and suggestions

The Yahoo! Web search exhibits similar characteristics with some top results arranged on a map. As an additional feature it provides a list of search terms to refine the search and also a list of suggestions while typing. These suggestions are generated from two sources, the n-grams and term relationships in matching documents and also previous search terms that have been entered by users in this combination. For location-based terms, these suggestions in most cases also provide relevant location-aware suggestions (Figure 12). An interesting detail is that although there are two cities of Oldenburg in the results, only the one with the larger population and larger amount of pages is featured in the map. This is similar to a disambiguation method proposed in (Markowetz et al., 2005). Bing Web search provides no discernible treatment of geospatial features.

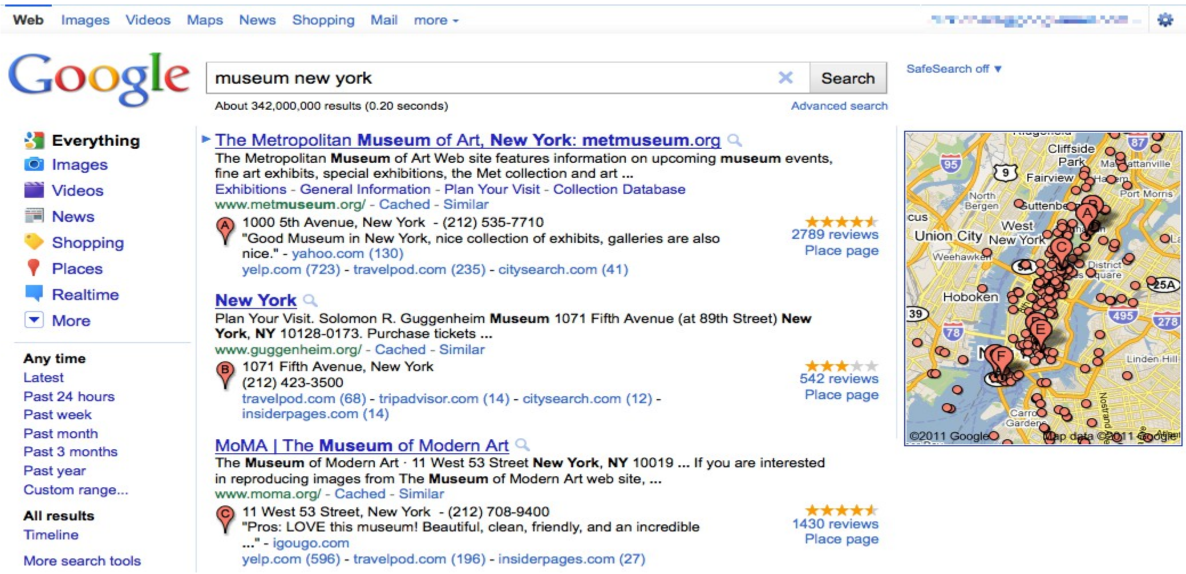

Fig. 13. Example of Google Place search in a museum search for New York

With the thorough integration of Web resources into Google Maps, an entry has Web pages associated with it and often the authoritative homepage of an entity. On the other hand, if the results of a general search return those homepages, they are then displayed as a general result list, but annotated with structured information from Maps and are accompanied by a miniature map view on the right side of the results, a view called Place Search[20] as seen in Figure 13. It also clusters results from different sources and is able to merge entities even when the name is slightly different. The interspersed results happen most often if a query asks for a keyword that has many homepage results, e.g., hotels or restaurants. Behind the scenes this means that for every general search, a local search is started as well and the results being integrated if they reach the threshold. On the other hand, for some local searches the top Web search result is also presented if it has an outstanding ranking. This leads to even deeper integration of local results into general Web search and sometimes removes the need for the user to even go to the separate local search.

In the Place Search, re-ranking and clustering happens very overt, with local results being styled differently, showing additional data and a red map pin marker. It can be observed that disambiguation happens in the place search. However, it happens in a way that prefers a larger place over a smaller one. If one searches for “museum oldenburg”, a general Web search returns results for both cities of Oldenburg that exist in Germany. However, the places search re-ranks the results so that results for Oldenburg in Oldenburg get the first result pages and results from the smaller Oldenburg in Holstein are placed completely on the second result page. A second, more hidden feature is the usage of GIR technology on Web pages that happens in the background and is used as one signal to influence the ranking and propagate potentially geographically relevant results to the top. The most obvious adaptation of search results concerns the language. If a language preference is given, the search is constrained to only documents of that language. A similar constraint is implicitly assumed based on domain names or more specifically, for a country-specific search, on the country-code top level domain (ccTLD) such as .de or .us, but also n the geographical server location. These play an important role in ranking documents, but since this is a mere technical features of the infrastructure rather than a content attribute, its influence is smaller than the used language. Furthermore, a

20 http://googleblog.blogspot.com/2010/10/place-search-faster-easier-way-to-find.html

geoparser tries to recognize toponyms in the page (similar to the Google Books example above). That is, documents are analyzed whether they discuss a certain region and will then be slightly higher ranked on respective queries containing a geographically similar concept or a nearby location. Another way this is used is in presenting the user with results nearer to his own location as another influence factor in ranking. In Place Search, especially high-ranking results of georeferenced results or a significant clustering of keywords in a certain region can trigger its use for a result page. Also for results that are otherwise rather homogeneously distributed this technique can help to select those results that distinguish themselves by being nearer to a searcher, in both desktop and mobile versions.

3.4. Other local search services

Some of the first Web systems embracing location were those where coordinate metadata could easily be added. Notable examples are flickr or wikipedia. The examples of the search engines that use reviews or even allow their users to make changes to the data shows the high influence of crowdsourcing in keeping databases current and complete.

Some services build themselves entirely around the concept of maintaining a database of user-generated reviews for places such as Yelp[21] or Qype[22]. As an add-on, locations appear in many other social network services, for example in Facebook Places. Using not only point locations, services such as foursquare[23], Yahoo FireEagle[24] or Google Latitude maintain traces of movement and build buddy finders or localized offers. But not only POI generation and placement can be crowdsourced, map generation can be as well. Applying the wiki concept to maps, OpenStreetMap[25] allows users to participate in map generation and management, including road network, buildings, and POIs.

Navigation systems mainly provide directions and routing calculations. Additionally, they carry a small POI database with them. While for most built-in systems today the database is still static, future systems will feature an online connection and be able to tap into the rich resources the Web and local search have to offer, bringing a more dynamic experience (Ahlers, Boll, Ebert, Fliegner, & Hofmann, 2008). A different visualization and access to geographic data is provided by virtual globes and mapping programs. For example, Google Earth draws from the same sources of Google Maps regarding current maps and entity search ability, but provides much more GIS-like functionality such as historic map data and the ability to include many different data sources as layers onto the globe. Similar services are offered as Nasa World Wind and Bing Maps 3D.

3.5. Impact

Similar to Web search engines before, local search engines are becoming the gatekeepers for local information. If before the saying went “You do not exist if you cannot be found on the Web” now it is “you do not exist if you cannot be located”.

As more people use local search and find it a convenient way to learn about their surroundings, plan a vacation or simply explore, it becomes vital to establish a searchable presence there as well. Who would reserve a hotel today that cannot be found on the Web, where you do not know its location and

21 http://www.yelp.com/

22 http://www.qype.com/

23 http://foursquare.com/

24 http://fireeagle.yahoo.net/

25 http://www.openstreetmap.org/

you cannot book online? While there are of course many people who still do that, in general, that number decreases. So it is not enough to just be visible on the Web, entities also have to be present in the geospatial Web and they have to make their services available. Additionally, depending on their line of business, it can be vital for them to be present in aggregator services as well. This will motivate page designers not only to make their page spatially searchable, but to provide as much information as possible in standardized formats to ensure an attractive listing.

An influence is also extended to search engine optimization (SEO) Apart from page optimization, the presence in local search results and the geographic targeting of an audience in both direct local search but also universal search is an issue, complemented by paid results and advertising. A spam problem exists also in this area, but the search engines seem to be able to fight it more efficiently. The other challenge is to ensure credibility and reliability of geospatial information that people start to rely on more and more.

The provision of APIs not only for the entity data and local search, but also for the map background data, has made cheap mapping tools available to developers. This provides the means for providers to rethink their geo strategy or simply makes new information and apps available, either as standalone services or in the form of geospatial mashups, thus opening up more data to the geospatial Web.

4. Mobile Local Search

Mobile search and local search are often used interchangeably. However, this is an oversimplification. Mobile search simply describes the search on a mobile device and in mobile queries, location plays an important role, but is surpassed by Web queries for similar topics as in desktop Web search. Study of mobile search behavior thus concerns also the type of queries, the length and other general features depending on the form factor of the devices (Amin et al., 2009; Kamvar & Baluja, 2006; Henrich & Lüdecke, 2007). But of course, location plays an important role to searches on mobile devices. The immediate surroundings of a user are often the focus of searches. In this case, the mobile device is specially suited to provide queries for local search. The device’s built-in GPS can detect the user’s location very exactly, making orientation easy by just having the map centered on that position.

The challenge in building a mobile local search solution is then to deal with the devices’ inherent limitations regarding size, input and output capabilities, battery life, computing power. The output can be presented on specially rendered maps, being easier to read on a mobile device, and can feature easily readable icons and representation of results on a map. As for input, the keyboards are much smaller, leading to significantly shorter queries (Kamvar & Baluja, 2006). However, as the location is already preselected or can easily be changed by the map interface, it actually eases interaction. Additional help can come from autocompletion of queries, not only based n general query log analysis, but also tailored to the current location, giving a better chance of guessing the right suggestions. In additionally helping the mobile user, speech recognition is being rolled out that further eases interaction with the search engine.

It should also be noted that mobile and/or local search does not always have to happen on a smartphone with a Web-based search or a custom application. More used in developing countries, phone or texting can also be an interface to a search engine. Projects exist that allow to query a search engine by a phone call, with speech recognition and synthesis happening on a server or also SMS-based systems that use a special vocabulary to exchange queries and abbreviated results to allow a search session. In all these cases, the mobile search accesses basically the same databases and indexes; only the interaction and presentation is adapted to the special use case and the available context.

5. Future Research and Evolution of Local Search

The increasingly spatial character of the Web and the inclusion of location into search and other applications as well as visualization, understanding and interaction will continue to change the way we use information from the Web. Already today, for many searches the cycle of query, result list, examination of results is moved over to a map-based interaction which allows first glimpses into possible interaction with spatial data. Easy to formulate, but difficult to formalize queries will enable new views into the data. Local search and GIR will help in geospatial decision making (Shahabi, Kashani, Khoshgozaran, & Xing, 2010) on a business level but also in the private life. For example, in GIR a query is represented as concept - spatial relationship - location. Today, only concept@location queries are supported. More powerful queries such as “a camping site near a river” will enable new views onto available data. A geospatial mining of locations and location trails of users can enable to find favorite places for individual users or for patterns within a city (Zheng, Zhang, Xie, & Ma, 2009). More context will be available to local search to better answer users’ queries not only about the current position, but also based on possible movement, motion patterns, and personal context (Ahlers & Boll, 2009). Much of the data needed for new scenarios is already on the Web, but is not yet made available to spatial search and reasoning. The challenges lie in the intelligent merging, aggregation and combination of spatial data to uncover hidden connections and to unlock the full geospatial potential of the Web.

References

Ahlers, D., & Boll, S. (2007). Location-based Web search. In A. Scharl & K. Tochterman (Eds.), The Geospatial Web. How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society. London: Springer.

Ahlers, D., & Boll, S. (2008). Retrieving Address-based Locations from the Web In GIR ’08: Proceedings of the 5th international workshop on geographic information retrieval. New York, NY, USA: ACM.

Ahlers, D., & Boll, S. (2009). Beyond Position – Spatial Context for Mobile Retrieval Systems. In Proceedings of the 6th workshop on positioning, navigation and communication 2009 (WPNC’09) (p. 129-134). Hanover, Germany.

Ahlers, D., Boll, S., Ebert, A., Fliegner, J., & Hofmann, M. (2008). Ortsbasierte Websuche im Automobil. In IMA’08. Brunswick, Germany.

Amin, A., Townsend, S., Ossenbruggen, J., & Hardman, L. (2009). Fancy a Drink in Canary Wharf?: A User Study on Location-Based Mobile Search. In Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I (pp. 736–749). Berlin, Heidelberg: Springer-Verlag.

Amitay, E., Har’El, N., Sivan, R., & Soffer, A. (2004). Web-a-Where: Geotagging Web Content. In Sigir ’04 (pp. 273–280). New York, NY, USA: ACM Press.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., et al. (2010). Google Street View: Capturing the World at Street Level. Computer, 43 ,32-38.

Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., & Raghavan, S. (2001). Searching the Web. ACM Transactions on Internet Technology, 1 , 2–43.

Arasu, A., & Garcia-Molina, H. (2003). Extracting Structured Data From Web Pages. In Sigmod ’03: Proceedings of the 2003 ACM sigmod international conference on management of data (pp. 337–348). New York, NY, USA: ACM.

Battelle, J. (2005). The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture. Portfolio.

Boll, S., & Ahlers, D. (2008). A Web more Geospatial: Insights into the Location Inside. In D. De Roure & W. Hall (Eds.), Workshop on understanding web evolution: A prerequisite for web science (webevolve2008) held at WWW08 (Vol. 300, pp. 40–43). Beijing, China: Web Science Research Initiative.

Broder, A. (2002). A Taxonomy of Web Search. SIGIR Forum, 36 (2), 3–10.

Bush, V. (1945). As We May Think. The Atlantic Monthly, 176 (1), 101–108.

Buyukkokten, O., Cho, J., Garcia-Molina, H., Gravano, L., & Shivakumar, N. (1999). Exploiting Geographical Location Information of Web Pages. In Proceedings of the ACM sigmod workshop on the web and databases (WEBDB’99) (pp. 91–96).

Cafarella, M. J., Madhavan, J., & Halevy, A. (2008). Web-Scale Extraction of Structured Data. SIGMOD Rec., 37 (4), 55–61.

Chang, C.-H., Kayed, M., Girgis, M. R., & Shaalan, K. F. (2006). A Survey of Web Information Extraction Systems. IEEE Transactions on Knowledge and Data Engineering, 18, 1411-1428.

Croft, B., Metzler, D., & Strohman, T. (2009). Search Engines: Information Retrieval in Practice (1st ed.). Addison Wesley.

De Sabbata, S., & Reichenbacher, T. (2010). A Probabilistic Model of Geographic Relevance. In GIR’10: Proceedings of the 6th workshop on geographic information retrieval (pp. 1–2). New York, NY, USA: ACM.

Downs, R. M., & Stea, D. (1977). Maps in Minds: Reflections on Cognitive Mapping. Publishers, Inc., 10 East 52rd Street, New York, N.Y. 10022, US: Harper and Row.

Ehlen, P., Zajac, R., & Rao, K. B. (2009). Location and Relevance. In Locweb ’09: Proceedings of the 2nd international workshop on location and the web (pp. 1–3). New York, NY, USA: ACM.

EUROSTAT Press Office. (2008). Use of information and communication technologies – STAT/08/173 (Tech. Rep.). EUROSTAT.

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., et al. (2001). What Makes Web Sites Credible?: A Report on a Large Quantitative Study. In Chi ’01: Proceedings of the sigchi conference on human factors in computing systems (pp. 61–68). New York, NY, USA: ACM.

Freedman, M., Vutukuru, M., Feamster, N., & Balakrishnan, H. (2005, October). Geographic Locality of IP Prefixes. In Internet measurement conference (imc) 2005. Berkeley, CA.

Gan, Q., Attenberg, J., Markowetz, A., & Suel, T. (2008). Analysis of geographic queries in a search engine log. In Locweb ’08: Proceedings of the first international workshop on location and the web (pp. 49–56). New York, NY, USA: ACM.

Goldberg, D. W. (2008). A Geocoding Best Practices Guide (Tech. Rep.). North American Association of Central Cancer Registries.

Gräf, J., Henrich, A., Lüdecke, V., & Schlieder, C. (2006). Geografisches Information Retrieval. Datenbank-Spektrum, 6 (18), 48–56.

Henrich, A., & Lüdecke, V. (2007). Characteristics of Geographic Information Needs. In GIR'07: Proceedings of the 4th acm workshop on geographical information retrieval (pp. 1–6). New York, NY, USA: ACM.

Heywood, I., Cornelius, S., & Carver, S. (2006). An Introduction to Geographical Information Systems (3rd ed.). Prentice Hall.

Hill, L. L. (2000). Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. In Ecdl '00: Proceedings of the 4th european conference on research and advanced technology for digital libraries (pp. 280–290). London, UK: Springer-Verlag.

Himmelstein, M. (2005). Local Search: The Internet Is the Yellow Pages. IEEE Computer, 38 (2), 26-34.

Hines, J., & Abou-Assaleh, T. (2009). Query GeoParser: A Spatial-Keyword Query Parser Using Regular Expressions. In Developer track, 18th international world wide web conference. ACM.

Jones, M. T., McClendon, B., Charaniya, A. P., & Ashbridge, M. (2007). Entity Display Priority in a Distributed Geographic Information System (US Patent 20070143345).

Jones, R., Zhang, W. V., Rey, B., Jhala, P., & Stipp, E. (2008). Geographic Intention and Modification in Web Search. International Journal of Geographical Information Science (IJGIS), 22 (3), 229–246.

Kamvar, M., & Baluja, S. (2006). A Large Scale Study of Wireless Search Behavior: Google Mobile Search. In Chi '06: Proceedings of the SIGCHI conference on human factors in computing systems (pp. 701–709). New York, NY, USA: ACM Press.

Larson, R. R. (1996). Geographic Information Retrieval and Spatial Browsing. GIS and Libraries: Patrons, Maps and Spatial Information, 81-124.

Lee, R., Kitayama, D., & Sumiya, K. (2008). Web-based Evidence Excavation to Explore the Authenticity of Local Events. In Wicow '08: Proceeding of the 2nd ACM workshop on information credibility on the web (pp. 63–66). New York, NY, USA: ACM.

Leidner, J. (2008). Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names. Universal Publishers. (PhD thesis)

Levene, M. (2006). An Introduction to Search Engines and Web Navigation. Addison Wesley, Pearson Education.

Lewandowski, D. (2009). Handbuch Internet-Suchmaschinen – Nutzerorientierung in Wissenschaft und Praxis. In D. Lewandowski (Ed.), (chap. Spezialsuchmaschinen). Heidelberg: Akademische Verlagsgesellschaft Aka.

Lewandowski, D., Nesbach, E., & Mikley, N. (2010). Lokale Suchmaschinen: Evaluierung und Gestaltungsempfehlungen. In J. Ziegler & A. Schmidt (Eds.), Mensch und Computer 2010 (p. 73-82). Oldenbourg Wissenschaftsverlag.

Lüdecke, V. (2010). Geographisches Information Retrieval für deutsche Webseiten – Problemstellungen und Lösungsansätze im Rahmen des Szenarios einer Deutschland-Suchmaschine. Doctoral dissertation, Otto-Friedrich-Universität Bamberg.

Lynch, K. (1960). The Image of the City. The MIT Press.

Marchionini, G. (2006). Exploratory Search: From Finding to Understanding. Commun. ACM, 49, 41–46.

Markowetz, A., Brinkhoff, T., & Seeger, B. (2004). Geographic Information Retrieval. In 3rd international workshop on web dynamics.

Markowetz, A., Chen, Y.-Y., Suel, T., Long, X., & Seeger, B. (2005). Design and Implementation of a Geographic Search Engine. In A. Doan, F. Neven, R. McCann, & G. J. Bex (Eds.),WebDB 2005 (p. 19-24). Baltimore, Maryland, USA.

Mountain, D., & Macfarlane, A. (2007). Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. J. Inf. Sci., 33 (5), 515–530.

Olston, C., & Najork, M. (2010). Web Crawling. Foundations and Trends in Information Retrieval, 4 (3), 175–246.

O'Sullivan, D., & Unwin, D. J. (2002). Geographic Information Analysis. Wiley.

Purves, R. S., Clough, P., Jones, C. B., Arampatzis, A., Bucher, B., Finch, D., et al. (2007). The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the internet [article]. International Journal of Geographical Information Science, 21 (7), 717-745.

Rattenbury, T., Good, N., & Naaman, M. (2007). Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. In SIGIR '07. ACM.

Rijsbergen, C. J. van. (1979). Information Retrieval (2nd ed.). London: Butterworths.

Scharl, A., & Tochterman, K. (Eds.). (2007). The Geospatial Web. How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society. London: Springer.

Serdyukov, P., Murdock, V., & Zwol, R. van. (2009). Placing Flickr Photos on a Map. In Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval (pp. 484–491). New York, NY, USA: ACM.

Shahabi, C., Kashani, F. B., Khoshgozaran, A., & Xing, S. (2010). GeoDec: A Framework to Effectively Visualize and Query Geospatial Data for Decision-Making. IEEE Multimedia.

Snavely, N., Seitz, S. M., & Szeliski, R. (2006). Photo Tourism: Exploring Photo Collections in 3D. In ACM transactions on graphics SIGGRAPH '06 proceedings (Vol. 25, pp. 835–846). New York, NY, USA: ACM.

Souza, L. A., Davis, C. A., Jr., Borges, K. A. V., Delboni, T. M., & Laender, A. H. F. (2005). The Role of Gazetteers in Geographic Knowledge Discovery on the Web. In La-web '05: Proceedings of the third latin american web congress (p. 157-165). Washington, DC, USA: IEEE Computer Society.

Steele, R. (2001). Techniques for Specialized Search Engines. In Proceedings of internet computing 2001. Las Vegas: CSREA Press.

Tang, J., & Sanderson, M. (2010). Spatial Diversity, Do Users Appreciate It? In GIR ’10: Proceedings of the 6th workshop on geographic information retrieval (pp. 22:1–22:2). New York, NY, USA: ACM.

Xiao, X., Wang, L., Xie, X., & Luo, Q. (2008). Discovering Co-located Queries in Geographic Search Logs. In Locweb ’08: Proceedings of the first international workshop on location and the web (p. 77-84). New York, NY, USA: ACM.

Zheng, Y., Zhang, L., Xie, X., & Ma, W.-Y. (2009). Mining Interesting Locations and Travel Sequences From GPS Trajectories. In J. Quemada, G. León, Y. S. Maarek, & W. Nejdl (Eds.), WWW 2009 (p. 791-800). ACM.