Technische Universität München, Lehrstuhl für Angewandte Informatik/ Kooperative Systeme

AN ARCHITECTURE FOR AN ALTERNATIVE, MULTI-AGENT-BASED INFORMATION RETRIEVAL APPROACH WITH SPATIO-TEMPORAL PRIMARY CLASSIFICATION CRITERION

Georg Groh, Florian Straub

Abstract: We discuss problems of current centralized Information Retrieval (IR)-approaches in view of a dynamic change in the culture of the Web towards a mobile, social Web of Participation where the paradigms of decentralization and autonomy play an increasingly important role. From the discussion of human principles of IR, existing approaches from Geo-IR, P2P-IR and the increased importance of spatio-temporal referencing of information, we justify the proposition of a distributed, Multi-Agent IR architecture with spatio-temporal reference as the primary classification criterion as an alternative solution to the future requirements in IR. We also discuss implicitly social and semantic spatio-temporal small world structures which play an important role in this concept. In a second part we discuss the elements of our architecture and their relations to the justifying theoretical considerations of the first part in more detail and present the processes in our approach as well as a qualitative evaluation with the help of a prototypical implementation of key aspects.

Keywords: Information Retrieval, Spatio-Temporal Information Retrieval, Distributed Information Retrieval, Social Information Retrieval, Multi-Agent-System, Distributed Spatio-Temporal Index, Peer-to-Peer-Protocol, Small World, Toblers First Law

// EINE ARCHITEKTUR FÜR EINEN ALTERNATIVEN MULTI-AGENTEN-BASIERTEN INFORMATION-RETRIEVAL-ANSATZ MIT RAUMZEITLICHEM BEZUG ALS PRIMÄREM ORDNUNGSKRITERIUM

// Zusammenfassung: Im Hinblick auf Beobachtungen bezüglich eines Kulturwandels im Web hin zu einem mobilen, sozialen Beteiligungs-Netz, in dem Autonomie und Dezentralisierung eine immer größere Rolle spielen, werden Probleme aktueller zentralisierter Information-Retrieval- (IR) -Ansätze diskutiert. Ausgehend von diesen Beobachtungen, den natürlichen kognitiven Ansätzen des Menschen bei der Informationsbeschaffung, existierenden Ansätzen aus Geo-IR und P2P-IR sowie der wachsenden Bedeutung der raumzeitlichen Verortung von Mensch und Information in einem solchen Netz beschäftigt sich der Beitrag mit einem alternativen Konzept einer verteilten Multi-Agenten-basierten IR-Architektur mit raumzeitlichem Bezug als primärem Ordnungskriterium. Neben einer Diskussion implizit sozialer und semantischer Small Worlds, die im vorgestellten Konzept eine wichtige Rolle spielen, werden im weiteren Verlauf des Beitrages die konstituierenden Elemente und Prozesse der vorgeschlagenen Architektur im Hinblick auf die zuvor diskutierten theoretischen Überlegungen genauer erläutert und abschließend eine qualitative Evaluation mit Hilfe einer prototypischen Implementierung zentraler Elemente vorgestellt.

Schlüsselwörter: Information Retrieval, raumzeitliches Information Retrieval, verteiltes Information Retrieval, soziales Information Retrieval, Multi-Agenten-System, verteilter raumzeitlicher Index, Peer-to-Peer-Protokoll, Small World, Toblers First Law

Anschrift der Autoren

Dr. rer. nat. Georg Groh
Technische Universität München
Lehrstuhl für Angewandte Informatik/ Kooperative Systeme
Boltzmannstr. 3
D-85748 Garching
E: grohg@in.tum.de

M.Sc. Florian Straub
Technische Universität München
Fachgebiet Geoinformationssysteme
Arcisstr. 21
D-80290 Muenchen
E: florian.straub@bv.tu-muenchen.de

1. INTRODUCTION

In view of future alternative Information Retrieval (IR) concepts, several important observations on current developments in the Web can be made which are supported by previous scientific work:

1. Autonomy and decentralization become increasingly important trends in the Social Web: Lowered publication hurdles result in an increasing amount of personal content exchanged among Social-Web prosumers. Because of the data silo character (Yeung 2009) of large centralized Social Networking platforms, this exchange must and will take place in an increasingly autonomous and decentralized fashion.
2. From the perspective of centralized search engines, the trend towards decentralized social networking increases the existing Hidden-Web-problem, because an increasing amount of personal information is going to be accessible for a selected audience only.
3. The Social Networking paradigm increasingly emphasizes a style of personal IR characterized by directly addressing other users from the social graph vicinity or querying their social web information spaces with queries that cannot as usefully be answered by using centralized search engines.
4. A new style of information publication on the Web is complementing the established ways of information publishing on the Web to an increasing extent. It is characterized by the change from anonymous, freely accessible, 1:n publication of *"cold"*, often factual, information with a globally comprehensible semantics for an unspecified audience to personal, access restricted, n:m communication (or intertwined, socially local information publication) of *"warm"* personal content in Social Networking and the Social Web with a more restricted contextual (e.g. spatio-temporal) scope and validity.
5. The Web is becoming mobile: Spatio-temporal context and other context-elements become increasingly important with respect to accessing services via mobile devices and also with respect to the content produced and communicated, because the mobile web is much deeper integrated into reality of people's lives.
6. Semantic small world structures (e.g. in networks of linked documents) and social small worlds (e.g. in friendship graphs) are related by spatial references of the network-nodes.

We will discuss and substantiate these observations in more detail in section (2).

Current centralized IR approaches are only partly able to deal with the implications of these observations:

- ad (1), (2): The Hidden-Web cannot be easily accessed by crawlers of centralized search engines. Specialized IR systems (e.g. a flight search engine on an airline's web page) have to take over in these cases.
- ad (3) (4): The described new style of information retrieval is not the primary focus of centralized search engines.
- ad (4), (5): If they can access it at all, centralized search engines have to cope with the local contextual semantics of the personal content communicated on the social web.
- ad (4), (5): Crawlers of centralized search engines can only access information periodically while in a mobile and social Web, information is produced and published asynchronously with an increasing dynamics and limited spatio-temporal scope and validity. This e.g. introduces a timeliness problem in centralized IR approaches.
- ad (6) While centralized search engines indirectly make use of semantic small world effects (e.g. via Page-Rank on document links) they do not fully exploit the connection between space, social relations and semantics.

While several attempts to overcome some of these problems exist (see e.g. (He 2007)) a general solution to the problems discussed above is still open. *"The one engine fits all model does not - cannot - scale well with the size, dynamics and heterogeneity of the web and its users"* (Wu 2007). In view of the observations, the problems they impose for centralized IR concepts (which will grow in importance over time) and the potential that these observations imply, we devote ourselves to the following *research question*:

How can architectures for alternative IR approaches be structured that better accommodate the given observations?

We will investigate this research question by pursuing a Design Science approach (Hevner 2004), which aims at developing *"a construct, a model, a method, or an instantiation"* (Hevner 2004) of Information Technology, innovatively contributing to the solution of a relevant problem with the goal of maximizing utility for users. In our case this implies constructing an architectural framework as an alternative IR approach in the aforementioned sense, grounding the approach on established previous work, providing detail solutions for occurring problems and finally evaluating key parts of the architecture with simulations and prototypical implementations, showing usefulness and generalizability of the approach.

In view of the research question, the observations discussed above imply characteristics and boundary conditions for such an alternative IR approach:

- A: Observations (1) and (2) imply that it is promising to construct and investigate decentralized approaches to IR. In a decentralized IR approach, agents autonomously control their local information spaces and provide locally optimized IR services. Distributed global index structures ensure that queries can be directed to some or all agents that can appropriately answer them. Thus the system of agents is collaborative on a basic level. Furthermore, agents can handle access policies on the level of queries or querying agents.
- B: Observations (3) and (4) also point to a decentralized system of autonomous agents, maintaining the personal information spaces of users in a decentralized social networking scenario. The resulting multi-agent system (together with the aforementioned distributed global structures) will contain agents of all sorts, purposes and sizes (from large airline information systems down to personal agents for individual persons).
- C: The aforementioned distributed global structures need to be balanced with respect to degree of detail and structuring criterion. A fully detailed distributed semantic index mimicking a centralized IR's index would inherit most of its problems discussed above. More coarse grained index structures which delegate specific levels of IR to the competent local IR agents are better able to deal with information dynamics and local

semantic validity. Observations (5) and (6) imply that spatio-temporal references may act as primary structuring criterion for distributed global structures (with semantic and social criteria as secondary criteria).

In this contribution we discuss an architecture for this alternative IR approach, whose design choices respect these aspects. The architecture is intended as a blueprint for systems who can deliver results different from what centralized IR approaches can deliver. It is thus difficult to compare these results in terms of precision and recall with the results of centralized approaches. So our goal is not to improve centralized approaches in the domain where they already perform well (e.g. retrieving factual information), but to rather complement them in the sense discussed above.

The remainder of the paper is organized in four sections. Section 2 lays a basis for our main argumentation and architectural proposal by discussing the observations from above in more detail. Section 3 investigates related work and elements of existing approaches from Geo-IR and P2P-IR. Section 4 presents our architecture in detail. We discuss how the observations discussed in section 2 influence the design of our architecture and describe the architecture down to a formal level. Section 5 presents a qualitative evaluation of the approach with the help of a prototypical implementation of key aspects. The conclusion (section 6) finally summarizes the contributions and gives an outlook on future research.

2. FOUNDATIONS FOR AN ALTERNATIVE APPROACH TO IR

In the following section, we will discuss and substantiate the observations which provide a foundation for our approach and which were shortly discussed in the introduction. We will do so by referencing the respective observations explicitly at the beginning of each sub-section.

Cf. Observation (2): An enormous number of documents (as containers of smaller information-items (as defined in (Groh 2005)) cannot be found or gathered by crawlers of centralized search engines at all (He 2007; Cho 2002). This phenomenon especially encompasses information-items in documents which, e.g. via Web-portals, Community- or Social-Networking-Sites, are either only available to authenticated and authorized users, which change frequently (see (Fetterly 2003; Ntoulas 2004)), or which are generated by request only (search-portals, flight-booking-portals, Web Services, Geo Web Services etc.). This problem is usually designated the Hidden-, Dark-, or Deep-Web-Problem (He 2007).

Cf. Observation (5): New classes of mobile devices like smart-phones or net-books allow for an ever deeper and smoother integration of the Web into our everyday lives and blur (and increasingly render meaningless) the borderlines between virtual and real world (see e.g. (Wellman 2004; Gartner 2010)). Mobile and ubiquitous Web-access does also change the Web itself, in that content published or searched and the services that operate on this content (like e.g. Geo-tagging-Services (Amitay 2004)) are increasingly sensitive to context, especially spatio-temporal context. As an example consider users creating a geo-tagged piece of content informing others about a traffic jam in a certain geographic region. This information will generally only be of interest for other drivers in the region.

Cf. Observation (1)(3)(4): The change to Web 2.0 is a change towards a Web of participation, where an ever increasing number of users contribute more content in a *"Peer-to-Peer fashion"*, thus effectively changing communication patterns from 1:n communication of conventional edited Web-sites to the n:m communication patterns of social network-platforms, micro-blogging-sites etc. where the user simultaneously acts as producer and consumer (*"prosumer"* (Toffler 1980)). Unfortunately, large Social Networking (SN) platforms such as Facebook or MySpace can, in general, not smoothly interoperate. They keep their users, networks and content locked in a data silo fashion (Yeung 2009). This implies decentralized SN with autonomous control over the data as an alternative. Furthermore, personal queries for information are often directed to friends in SN platforms instead of querying centralized anonymous search engines (Nielsen 2009), because they possess information which can be more context-relevant and is often not accessible to centralized search engines at all. A *"situative"* and *"situatively used"*, dynamic and personal Web of such kind requires new IR approaches in order to satisfy personal information needs, where "situative" implies that (mobile) Web-applications can sense and incorporate individual and social contexts and *"situatively used"* implies that e.g. via mobile interaction new types or references of content are produced. As an example for the above observations, consider a user asking for good restaurants in a friend's hometown that she pays a visit to. She may put more hope into asking her friend than doing research on restaurants with the help of centralized IR.

Cf. Observation (3)(4): In view of alternative approaches for IR in an increasingly mobile and social Web, it is reasonable to investigate and respect natural behavior and processes of human information gathering. This human IR principle where space and time play an important role will be a basis for our architecture. In order to define this principle, we will characterize the way humans situated in space and time and in a social network accomplish information retrieval using this social network, spatio-temporal references of information and humans, as well as topical references. As an information seeker (consumer), successful human information retrieval often boils down to finding the right information producer.

The human cognitive process of searching for information usually first implies asking yourself WHAT exactly to find (specifying your topical information need) and usually also implies asking yourself WHO (the right person, the person that holds relevant information) to ask for this information. Another component is usually WHERE information might be located, which is connected to the question of WHERE the aforementioned source (WHO) is staying and also how to access this information via this source. This is a first function or semantic variety of a WHERE specification. In case of spatially related information, WHERE (in a second function or semantic variety) also refers to the explicit or implicit spatial relation of the information itself (e.g. that the information is about a certain place). In connection with WHO, this second variety also implies estimating WHERE a possible expert might have her spatial center of expertise. Another important specification is obviously WHEN, specifying the temporal relatedness of information. WHEN (in a similar fashion as WHERE) also comes in different semantic varieties and functions. In contrast to WHERE, the temporal position of a person with suitable WHAT expertise in view of accessing the desired information is usually

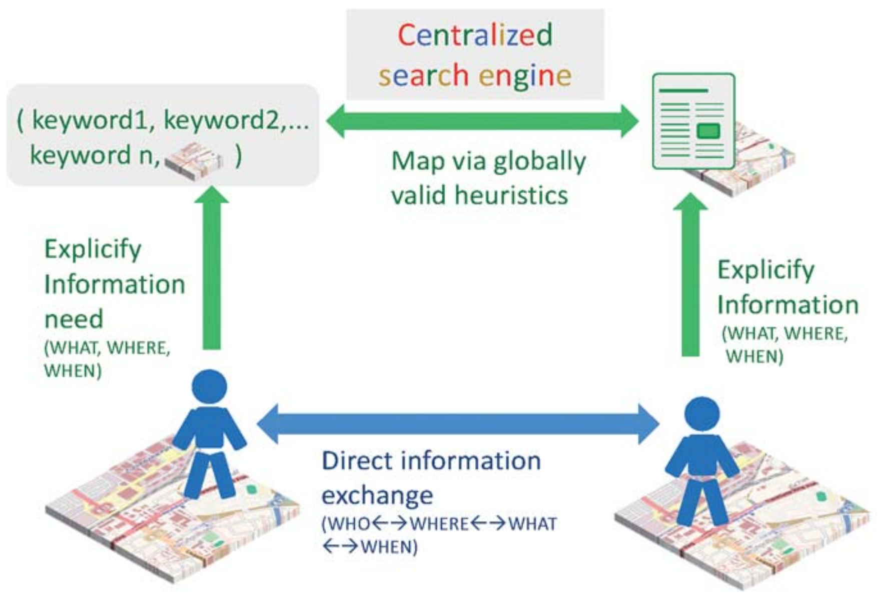

Figure 1: Global (Geographic) IR vs. Local (Geographic) IR.

not considered, because position in time cannot be arbitrarily chosen and is confined to the present. However, temporal relatedness of information and human expertise with respect to certain time periods is similar to the spatial case.

For evaluating the involved, interrelated WHO < > WHERE < > WHAT < > WHEN-specifications, humans rely on subtle previous knowledge concerning these specifications, which may not necessarily be explicitly reflected in the query that results from the human information need. As an example consider asking a friend for tickets to *"the"* ACDC concert. The query will be understood without specifying that the next concert taking place in Munich is meant because the friend shares the spatio-temporal context with the asking person.

These fine grained and subtle reasoning mechanisms in human cognition which define the human IR principle are not easily modeled in queries for a centralized IR system. Because for *"anonymous"* information publishing and subsequent retrieval on the Web, human-bound information and its underlying and implicit WHO < > WHERE < > WHAT < > WHEN-specifications have to be made explicit by coding it into text or even formal semantic metadata (Antoniou 2008), thus performing an explicification mapping. This explicified information can then be indexed and made searchable by centralized search engines (see figure 1). A centralized search engine has the advantage of making the WHO and the first function and semantic variety of WHERE specifications of human IR queries transparent as a first order approximation, because the engine takes over or directly mediates the role of WHO and WHERE.

However, the explicification mapping on the information producer's side, which naturally includes semantic losses, and the related IR heuristics on the search engine's side need to have global significance. In order to accomplish e.g. a search for a pizzeria in Munich, the centralized engine has to develop globally valid heuristics that e.g. model the WHERE < > WHAT relation. Furthermore, a user would have to explicitly specify Munich because the engine cannot automatically guess the context of a query.

In a system of distributed, collaborative IR agents with the capability to fully handle, process, compare, produce and manipulate spatio-temporal specifications, those agents can in a much easier way use shared contexts to skip certain explicification mappings and also use partial explicifications with local meaning only. Local IR systems are free to treat specialized problems in connection with spatio-temporal references or spatio-temporal context respectively, in a locally optimal way, depending on the intended focus of the local IR system. Among these problems are the differentiation between the geographic reference (Location) and relevance (Locality) (Gräf 2006) of information or queries or the intricate ways in which a human's perception of WHERE may differ from a mathematical or geographical representation of WHERE (Egenhofer 1995; Gluck 1995). Thus in such a distributed system of spatio-temporally enabled IR agents there is less need for a globally valid set of respective spatio-temporal heuristics.

Cf. Observation (5): Spatial reference as context in general information retrieval has always been important (Jones 2008; Larson 1996) and will gain importance in the future (Mountain 2007). A reason for this growing importance is that it can reasonably be assumed that mobile interacting with IR systems will more often induce an information need with spatial reference than interacting with IR systems in a desktop scenario. According to (Albaredes 1992), about 80 % of all decisions of an individual are related to a direct or indirect spatial reference. According to (Sanderson 2004), about 20 % of all Web-searches (not distinguishing between mobile and desktop access) have a geographic context. In a mobile and social Web, information is produced and published with limited spatio-temporal scope and validity.

Temporal reference and temporal context is also of great importance in IR, not just with respect to explicit or implicit temporal references in the queries or information-items themselves (as in the mobile interaction scenario just discussed) but also with respect to timely accessibility of time critical information. Crawlers of large centralized IR systems systematically have some inertia (Bawa 2003) while in a mobile and social Web, information is produced and published asynchronously with an increasing dynamics. This e.g. introduces a timeliness problem in centralized IR approaches. If the information has limited temporal validity this may be a severe drawback. Thus, instead of pull-based, synchronous crawler-style information access, a push based asynchronous producer-driven approach would be necessary to cope with the timeliness problem, which is difficult to enforce. It is thus an upcoming trend to consider real-time (or almost real time) information retrieval approaches in centralized IR systems, Isolated deals (see Abell 2009) between large centralized search engines and micro-blogging sites show that the industry is becoming aware of the problem. However, until now such arrangements will only be able to implement information streams with nearly real time quality between central IR engines and large content providers or, in general, those information sources that are willing or able to provide such push based information exchange. The fact that such compromises are established shows the problem of traditional central IR architectures with respect to real-time IR. An alternative system of distributed IR agents is better able to deal with this problem because no full central index of information is necessary.

Cf. Observation (6): A small world is a structural type of graph that is characterized by dense sub-networks (thus having a characteristically large clustering coefficient) which in turn are interconnected by a number of inter-cluster edges and thus exhibits a characteristically small diameter (Milgram 1967; Watts 1998; Kleinberg 2000a). Small world structures can regularly be found in a lot of naturally occurring networks. E.g. it is known that human social networks have a small world structure (see (Kleinberg 2000a)). Furthermore, (see e.g. (Newman 2006)) the link-graph of the documents on the Web has also been shown to exhibit small world structure. An important insight gained from the experiments of Milgram (Milgram 1967) and subsequent experiments is that in small worlds routing without global knowledge of the network but with local knowledge only is possible (Kleinberg2004). For such a routing (decentralized search), an understanding or measure of distance (e.g. social, semantic or geographical distance) needs to be present (Kleinberg 2000b; Liben-Nowell 2005).

In case of a human social network, nodes of the network represent persons and edges represent social relations between those actors. Formation of clusters in such a structure occurs on the basis of social distance (path-length between two nodes in the graph). If the semantics of nodes, edges are of social nature and clusters have a social significance, we can speak of a (purely) social or social-topology small world. In case of the link-graph of documents on the Web, nodes represent documents and edges represent semantically or topically relevant links between documents. Clustering may happen on the basis of semantic distance (path-length between two documents). Thus, semantics of nodes and edges are of semantic nature and clusters have a semantic significance. In this case we can speak of a (purely) semantic or semantic-topology small world.

Centralized search engines only indirectly make use of semantic small world effects (Kleinberg 1998). We will show that exploiting such small worlds denoting a small world graph with a real world geodetic reference frame (or spatio-temporal reference frame, respectively) as a spatial small world (or spatio-temporal small world, respectively), spatio-temporal small worlds can be of implicitly social or implicitly semantic nature, conserving their key topological properties: Clusters resulting from spatio-temporal locatedness will imply clusters in the corresponding social and / or semantic small world, which we will use for IR.

3. RELATED WORK AND ELEMENTS OF EXISTING APPROACHES

Motivated by the observation that spatial and temporal references are of great importance for our alternative IR approach, we will now shortly discuss existing approaches in IR to treat space and (to a lesser extent) time.

Geographic IR (GIR) is basically concerned with retrieval of information from geo-referenced information resources resulting in spatial indexing, retrieval and ranking (Jones 2008). Standard IR systems are augmented with components that e.g. explicify implicit spatial references of queries or information-items, process queries with geographic context or are able to present and visualize the spatial references of the results in a suitable manner (Vestavik 2004). The explicification of the implicit geographic reference of an information-item or a query can be accomplished by geo-parsing (Jones 2008; Larson 1996). This approach uses extensions and adapted versions of Named Entity Recognition (NER) from Natural Language Processing which are often tailored towards entities of type *"place"*. The explicified geographic reference is then transformed by geo-coding (Larson 1996) into a geographic footprint (Hill 2000). Having identified and modeled spatial references of an information-item, an indexing schema allows the system to retrieve spatially related information in an efficient way (Andrade 2006; Bliujut'e 1998). In GIR, usual techniques for indexing with respect to non-spatial content of an item (e.g. inverted lists) are combined with a spatial index (Jones 2008; Vaid 2005) like Quadtrees (Samet 1984) or R-Trees (Guttman 1984; Beckmann 1990; Saltenis 1999).

In GIR, a query is usually modeled as a triple <topic keywords, spatial relation, location>, combining content specification with a location and a spatial preposition specifying a spatial relationship connecting the non-geographic aspect of the information need with the location (like *"north of"*, *"near by"*, etc.) (Purves 2006). The combination of location and spatial preposition induces a geographic footprint that is matched against the geographic footprint of the information-item using spatial heuristics. E.g. Euclidian distances in a coordinate space of the overlap of minimum bounding rectangles (MBR) of two regions are used as a measure of spatial relevance (Larson 2004). Documents having been selected in that way are thus relevant with respect to geographic relation as well as topical relation. In view of relevance-ranking of documents, both relevance aspects have to be combined.

Motivated by the observation that distributed systems of IR agents are interesting alternatives for future IR systems, we will now briefly discuss existing Peer-to-Peer (P2P) approaches as realizations of communication-protocols in such distributed Multi-Agent Systems (MAS) and existing approaches from P2P-Information retrieval (P2P-IR).

The edges in a P2P network's graph represent routing information in the sense of *"knows"* relations. Systems of autonomous agents in a MAS may act on top of and use P2P-approaches e.g. for communication. In contrast to centralized IR, in P2P-IR , peers holding an information-item (or more generally peers able to process a query) have to be found in a collaborative and decentralized way (Steinmetz 2004). Several approaches for efficiently routing a query to just those peers that may hold potentially relevant items have been developed by the research communities over the last years (e.g. see (Lau 2005; Risson 2004)). One example are semantic topologies which use the semantics of the information in the information spaces of the peers for constructing topologies (Haase 2004) and thus for routing (Joseph 2002). The P2P network is structured according to topical relations and so, peers with similar information form semantic clusters in the network which can be used for an efficient query processing (Risson 2004; Löser 2005; Tang 2003). Furthermore, the small-world topology often found in networks structured according to social or semantic principles contributes to and allows for efficient decentralized search (Kleinberg 2004) e.g. used in the approaches of (Schmitz 2005) and (Li 2004).

Basic P2P-IR approaches, like distributing full indices, e.g. with distributed hash tales (DHT), alone does not automatically provide an alternative IR approach in our sense: In essence, the simple model of semantic relatedness of a query and an information-item that is implemented by compiling word- or n-gram-based full indices (Manning 2008) of the respective documents and comparing the query against these indices still prevails and still remains only a rather primitive approximation of semantic relatedness. The results will, in general, not be better qualitatively than those of a centralized index. In fact, a centralized search engine can employ a number of improvements to the approximation of semantic relatedness which are not easily mapped to a DHT based P2P-IR approach.

For a true alternative IR concept which stays abreast of changes with respect to an increasingly distributed social and mobile Web in an adequate way, peers must become competent IR agents, which can autonomously act as individual experts with deep access and expertise in confined information spaces and with the ability to exchange and compare elements of these information spaces with other agents.

4. ARCHITECTURE FOR AN ALTERNATIVE MULTI-AGENT-BASED IR APPROACH WITH SPATIO-TEMPORAL REFERENCES AS PRIMARY CLASSIFICATION CRITERION

With respect to realizing this approach, it is reasonable to lean on principles of natural behavior and processes of humans in information gathering (human IR principle) using explicit and implicit WHO < > WHERE < > WHAT < > WHEN-specifications to find other users that provide the right information with special regard to spatio-temporal reference of persons and information.

We thus propose an architecture of autonomous IR agents, each controlling its own local information space and providing access to it via appropriate IR services. The agents collaborate via adapted P2P approaches using a special form of a distributed global index that allows for locating agents with a high probability of being able to answer queries appropriately. The index acts as a common knowledge base for the agents and uses spatio-temporal references as primary classification criterion respecting the abundant location of humans and information in space and time. It is crucial that it has the right granularity: While the timeliness problem of centralized IR could in principle be slightly alleviated by using a distributed full index of all information-items and have the agents update it asynchronously, it would still be difficult to employ elements such as appropriate access control and use the expertise of local IR systems of agents. We will thus use a more coarse grained distributed index. We naturally include a decentralized Social Networking principle, by delegating to the agents decisions of suitability of information with respect to a query, of access control for information and of potentially mediating and establishing direct communication channels between information seeker and potential information providers. The approach uses small world effects relating space, time, content-semantics and persons as information providers, in order to deliver new forms of IR results, e.g. such results that exhibit spatial, semantic or social relations to the query that have not explicitly been searched for.

4.1 SMALL WORLDS AND SPATIO-TEMPORAL REFERENCE AS KEY STRUCTURING ELEMENT

In our approach, we argue towards a paradigm of spatio-temporal referencing (WHERE+WHEN) of Information-Items (whose topic corresponds to WHAT) and of users (WHO) as the primary classification criterion optionally supplemented by other (e.g. semantic / topical) classification criteria as sub-ordinated criteria.

With respect to that assumption, so called Tobler's First Law of Geography (TFL) formulated in 1970 by the geographer Waldo Tobler plays an important role. This law states that *"everything is related to everything else, but near things are more related than distant things."* (Tobler 1970)

TFL is predominantly applied in relation with spatially continuous phenomena in geo-information-science. Not only is Inverse Distance Weighting (IDW) as special instance of distance based spatial interpolation based on TFL, but also many geostatistic methods such as Kriging. Applying TFL to graphs modeling a social or semantic network by locating the nodes in a space with underlying geodetic reference system implies the implicit conservation of the structural social or semantic small world properties of these graphs in that space (there will be outliers of course): Introducing a geodetic reference frame to nodes of a social or semantic network does lead to a change in the overall structure regarding the introduced geodetic reference frame, but substantial characteristics of the original topology and thus the social or semantic context of the original graph can still be found in that space, because entities being geographically near are, according to TFL, also statistically more socially and semantically related. E.g. in a semantic small world structure as discussed before this means that according to TFL, it can be expected that for three documents A, B and C with spatial distance $d_{spatial}(A,B) < d_{spatial}(A,C)$, document A and B have a higher probability of (directly or indirectly) linking to each other than A and C implying that the semantic distance $d_{semantic}(A,B) < d_{semantic}(A,C)$. Here, the spatial reference of a document may e.g. be computed as the centroid of the spatial references of all entities (places, persons etc.) that the document refers to.

In view of social and semantic context, TFL was just recently justified and empirically verified by an extensive study by (Hecht 2009) using Wikipedia articles together with their link and author structure. Thus in a social or semantic network with an according small world structure, where nodes are located in a real-world geodetic reference frame, cluster formation occurs (and routing can be accomplished) (see (Newman 2006) and the work by Milgram (Milgram 1967)) based on spatial distance between nodes and maintains the social and semantic cluster-structure in a statistically significant way (Hecht 2009). These observations led to an idea stated as First Law of Cognitive Geography: *"people believe closer things are more similar"* (Fabrikant 2002; Montello 2003). The same argumentation also applies to temporal reference frames and also to combinations, that is spatio-temporal reference frames (see (Hecht 2008b)).

We will denote a small world graph with a real world geodetic reference frame (or spatio-temporal reference frame, respectively) as a spatial small world (or spatio-temporal small world, respectively). Thus, depending on the original social or semantic topology, spatio-temporal small worlds can be of implicitly social or implicity semantic nature (links (2) and (3) in figure 2).

As an example, a spatio-temporal location A implies that (a) if some friends are spatially related to A (e.g. have their centers of life there) then it is quite likely that they are friends and (b) documents whose (explicit or implicit) spatio-temporal reference is A are quite likely to link to each other (or have similar content). With respect to e.g. (a), implicitly social spatio-temporal small world implies that in the example there will also be some friendship relations from the friends *"clique" "at"* A (e.g. a small village), pointing to people located at far away from A (e.g. some friends in a larger town) (Liben-Nowell 2005). Statement (a) can also be justified by socio-psychological means (see e.g. (Groh 2005) and especially (Nahemov 1975)).

The link between social small worlds and semantic small worlds (link (1) in figure 2) and the spatio-temporal locatedness of nodes (links (4) and (5)) in both types of small worlds is established by stating that the documents in an information space of a human with spatio-temporal location A (center of life) can be expected to also significantly often have an (explicit or implicit) spatio-temporal reference to A (see figure 3). This is reasonable because it can be assumed that most people's interests are distributed like that (Lieberman 2009). Furthermore, direct studies have also given evidence to that statement (see e.g. (Hecht 2010a)). Thus, spatio-temporal reference acts as a linker between social small worlds and semantic small worlds (as visualized in figure 3). Another observation with respect to the mixture of social and semantic aspects is that e.g. in a social small world, the inherently social nodes (e.g. via node profiles) can also carry semantic information (e.g. interests of the person) or spatio-temporal information (e.g. the person's current location).

So node profiles can be used for routing and clustering in this small world: In contrast to other approaches in IR (e.g. see (Haase 2004; Schmitz 2005; Li 2004)) that actively use semantic clustering of documents for supporting the IR process, our approach just relies on the topically spatio-temporal location of Information-Items or documents, because the clusters resulting from spatio-temporal locatedness will, as argued above, also imply clusters in the corresponding social and / or semantic small world (within the limits of the statistical tendency characterized by TFL (Hecht 2009)), which we will use for IR. Introducing a spatio-temporal reference frame as a primary classification criterion and the related implicit social and semantic structures have effects in our concept at various levels which we will discuss below.

If we transfer the linking from the level of individual Information-Items (inducing a semantic small world) to the level of confederated IR agents by having each agent maintain a list of *"Expert-Links"* to other agents (inducing a social small world), specifying WHO the other agent is, WHAT he is expert for (semantic small world) and what the spatio-temporal reference of that information-competence is (WHERE+WHEN), we have a spatio-temporal small world on the level of Information-Items with an implicit semantic structure and on the level of IR agents with an implicit social and semantic structure. Since interests govern information needs

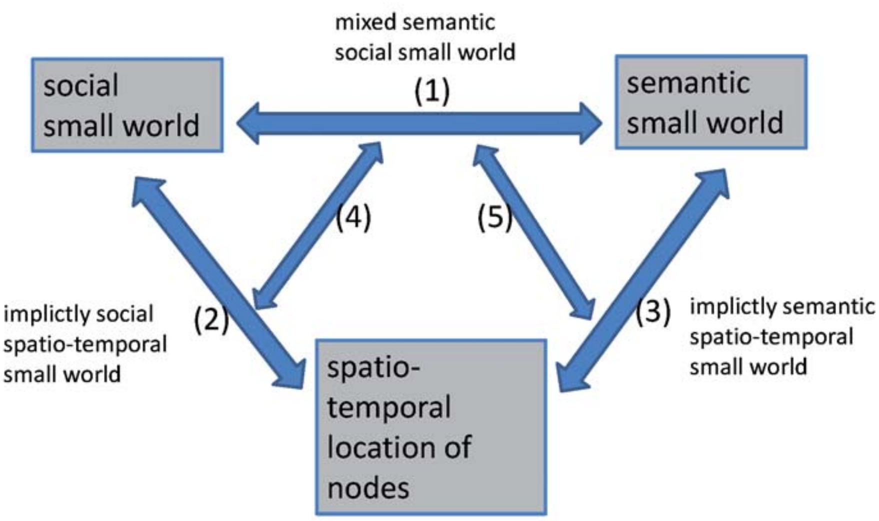

Figure 2: *Links between social small worlds and semantic small worlds and the spatio-temporal locatedness*

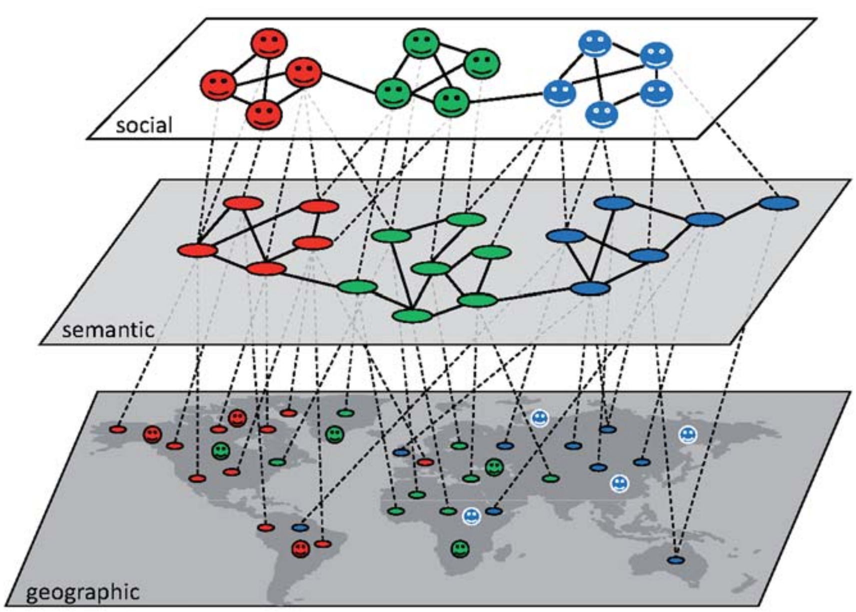

Figure 3: *Preserving and Connecting Social and Semantic Small Worlds via Spatio-Temporal Referencing*

and information needs imply Information-Items found by asking other agents, the Expert-Links pointing to other agent's Information-Items should also exhibit the spatio-temporal small world structure.

For referencing our key idea and for enabling intelligent collaboration between agents, in the style of the natural attitude of humans and the resulting intuitive processes with respect to basic physical facts of space and time in information gathering, we need to transfer these ideas to our conceptual architecture, which we will discuss in the following.

4.2 CONCEPTUAL ARCHITECTURE OF THE MULTI-AGENT-SYSTEM (MAS)

Every agent in the system acts like a person in a social network. It is autonomous, can join or leave the network at will and has a own local information space. It learns by user triggered (or agent triggered) querying of other agents and evaluating this exchange with other agents by establishing or modifying Expert-Links. Expert-Links thus represent knowledge about other agent's knowledge. Having an information need, it chooses other agents to ask on the basis of own experience (Expert-Links) and on the basis of research in the distributed and spatio-temporally organized global knowledge base (the global index). In this index, each agent deposits his current position, abstracts of his own knowledge in form of Expertises, and Expert-Links to other agents. By querying this index with a spatio-temporal query, an agent receives all agents that have self declared Expertise or expertise attested by other agents with respect to that region or which have a current position there. Via TFL an agent thus not only uses the spatio-temporal small world structure of the network but implicitly also the social and semantic small world structure existing there.

The MAS contains four groups of entities: agents, queries, answers to queries and the distributed spatio-temporal index (DSTI) and is organized as a P2P network. Queries are employed to find agents offering required Information-Items within the network. Agents may answer the queries.

MAS = (AGENTS, QUERIES, ANSWERS, DSTI)

The distributed spatio-temporal index (DSTI) is a three dimensional spatial index-structure on the basis of a distributed Quadtree. In the style of other structured P2P protocols, certain peers in the system are responsible for certain sub-trees and keep these sub-trees. Comparable examples for index-structures are e.g. P2PR-Trees (Mondal 2004) or the approach described in (Tanin 2007) based on a Quadtree index structure. In the style of other distributed hash-table based approaches for P2P-Systems like CAN, the mapping from peers to resources is accomplished via hashing peer-addresses and content-keys, in our case WHERE+WHEN as 3D coordinates (x,y,t) into a common key-space. Because of the fact that it is a spatial index, the identity function (h(x,y,t) = (x,y,t)) can be used as hash-function.

Following the spirit of the definition in (Groh 2005), an Information-Item is the smallest entity in the information space of a peer which can be assigned (WHAT, WHERE, WHEN)-meta-data where WHAT defines the topic of the Information-Item and WHERE+WHEN (ST-LOCATION) defines the location in space and time which the Information-Item explicitly or implicitly refers to. A document in the information space may contain one or several Information-Items. Each Information-Item contains a link to its document. Indexed Information-Items represent the aforementioned (WHAT, WHERE, WHEN)-meta-data in a form suitable for use in the local IR System of the peer. All Information-Items of all peers in the entire network form the global community information space (CIS) of the System.

INDEXEDINFORMATIONITEM = (WHAT, ST-LOCATION)

ST-LOCATION = (WHERE, WHEN)

WHERE is split into a tuple <S-RELATION, LOCATION>, where LOCATION specifies the geographic areas of interest and S-RELATION specifies a spatial relationship connecting WHAT (the non-geographic aspect of the information need) with the LOCATION.

WHERE = (S-RELATION, LOCATION)

S-RELATION = (e.g. 'west of', 'near' or 'within 3km')

LOCATION = (POINT | LINE | REGION)

WHEN is also split into a tuple <T-RELATION, TIME>, where TIME specifies the temporal specification of the information need and T-RELATION specifies a temporal relationship connecting WHAT (the non-temporal aspect of the information need) and the TIME.

WHEN = (T-RELATION, TIME)

T-RELATION = (e.g. 'before', 'after' or 'between')

TIME = (T-POINT, T-INTERVAL)

An agent in our system has five subcomponents called spatio-temporal sub-system (ST-SUBSYS), local information space (LI-SPACE), local spatio-temporal IR system (STIR-SYSTEM), set of Expertise-Cards and set of Expert-Links. Thus a peer agent can be represented as a five-tuple

AGENT = (ST-SUBSYS, LISPACE, STIR-SYSTEM, EXPERTISE, EXPERT-LINKS)

Every agent owns its own spatio-temporal sub-system (ST-SUBSYSTEM) which means it is fully spatio-temporally-enabled and therefore always aware of its spatio-temporal position (e.g. using GPS). The local information space serves as a storage space for all information-items / documents an agent owns. On top of this a local (spatio-temporal) information retrieval system (STIR-SYSTEM) not only allows general handling of spatio-temporal references, comparable to a geographic information system (GIS), and (spatio-temporal) information retrieval over the local information space but also acts as an interface to the global information space.

The set of Expertise-Cards (the *"Expertise"*) is computed from those indexed Information-Items that a peer considers for public access. An Expertise-Card references a cluster of Information-Items from the information space of a peer, which is the result of a spatio-temporal clustering. In essence, an Expertise-Card consists of a spatio-temporal location (WHERE+WHEN), suitably computed from the very similar WHERE+WHEN specifications of the indexed Information-Items it consists of (more specifically the minimum bounding cuboid (MBC) containing all WHERE+WHEN specs of these (indexed) Information-Items), and a reference to the owner of the Expertise-Card. As discussed abouve, Expertise-Cards may also be manually created.

The approach of using the spatio-temporal location of Information-Items as the primary classification criterion, follows from the discussion in section 2, especially based on Tobler's first law (Tobler 1967) and the observation in (Hecht 2009) and thus from the fact, that semantic relatedness of information can be expected to be indirectly proportional to the distance of their spatio-temporal focus with high probability, thus forming an implicit semantic small world.

While the Expertise represents a part of the locally indexed contents of the personal information space of a peer, Expert-Links represent links to other agent's competencies which either have been manually added or are related to past queries successfully answered by Information-Items from those agent's information spaces or to queries from other agents concerning a certain topic. Expert-Links are maintained in relevance ranked list (Expert-List) specifying the quality of the respective expertise of a foreign agent pointed to. The Expert-List represents an agent's experiences concerning past retrieval processes. Agent's constantly delivering good answers are automatically ranked higher. This ranking in the list (which is part of the Expert-Link) can be used to e.g. estimate the expected quality of a delivered answer and thus represents an expression of trust, credibility etc. An agent leans by updating this list.

Expert-Links to other agents / peers represent the edges between peers of our system in our small world structure on the level of peers. An Expert-Link essentially consists of an index representing the position in the Expert-List of the peer having created the Expert-Link, a (WHERE, WHAT, WHEN)-specification, specifying the competence of the target peer (e.g. acquired through the original query that led to the establishment of the Expert-Link) and a link to that peer. Such a specification need not be complete (e.g. because the Expert-Link was manually created and incomplete, because some elements were unknown). Manual creation may represent personal experiences of the person behind the agent / peer.

Because Expert-Links are links between agents, agents plus Expert-Links can form an explicit social small world and, via spatio-temporal referencing of the nodes, can also form an implicitly social spatio-temporal small world as defined in section 4.1. More specifically, Expert-Links link to specific expertises of agents. Since the edges carry a (WHERE+WHAT+WHEN) specification (which is inherently semantic), and since the respective Expertises also obviously carry this semantics, Expertises of all agents plus Expert-Links also can form an (implicit or explicit) semantic small world and, analogously to above, via spatio-temporal referencing of the nodes, can also form an implicitly semantic spatio-temporal small world.

Expertise (set of Expertise-Cards) and Expert-List (ranked list of Expert-Links) together represent the overall indexed to-be-published knowledge of an agent (published as Knowledge-Flags (see below)).It should be emphasized that the overall published indexed knowledge of an agent does not need to be a mapping of the whole information in the local information space of an agent. This can have several reasons, for example the decision what to publish in this form is made by the person who controls the agent (e.g. manually, on the basis of appropriate rules). This does not limit the agent's subsequent decisions to further control actual access to the information itself e.g. on the basis of social aspects). Another reason could be that not all information (or knowledge or expertise) of a user must be present in his local information space. But just that personal knowledge that is not explicitly represented and the control over that knowledge and the related social aspects are an important aspect of our architecture (human IR as a role model). Including the user, the system in that way allows controlled access to "warm" personal information of a person behind an agent by supporting IR related human to human communication and thus by using the social small world structures present in the system.

A Knowledge-Flag represents the instantiated version of an Expertise-Card or an Expert-Link on the DSTI. From the view of the whole system, a Knowledge-Flag is a version of (from the view of an agent) local Expertise-Card or Expert-Link plus the unique ID of the peer that introduced the flag into the system. Thus a Knowledge-Flag originating from an Expertise-Card is essentially of the form (Agent A, (WHERE + WHAT + WHEN), Agent A) and a flag originating from an Expert-Link is essentially of the form (Agent A, (WHERE + WHAT + WHEN), Agent B).

Each query is a tuple < WHAT, ST-LOCATION >, where WHAT specifies the non-spatio-temporal aspect of the information need and ST-LOCATION specifies the spatio-temporal area of interest. Thus the formal definition of a query is analogous to the formal definition of an indexed Information-Item. With respect to the explicit presence of WHERE, WHAT and WHEN the following degrees of freedom are supported:

(WHAT | NULL) & ((WHERE & WHEN) | (WHERE | WHEN | NULL))

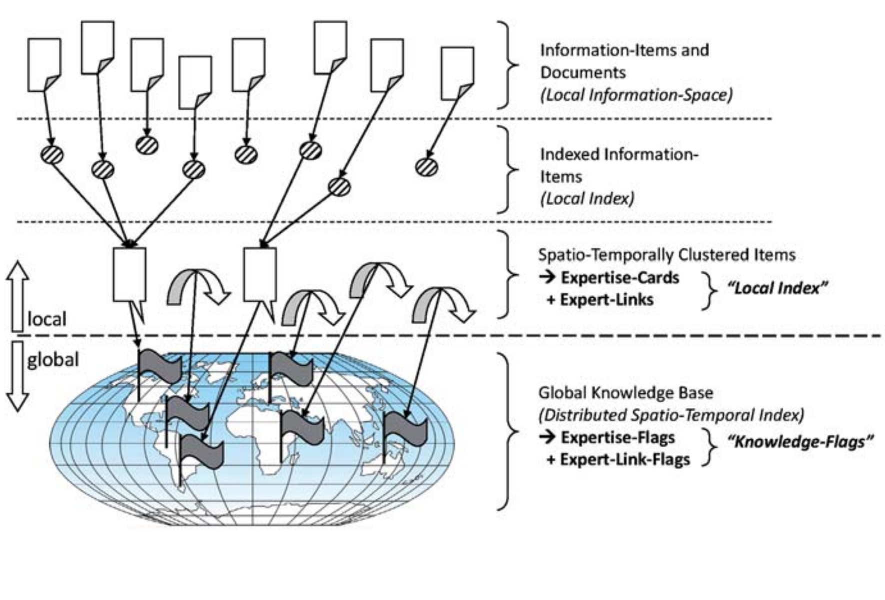

Figure 4: Spatio-Temporal Referencing: From Local Information Space to Published Knowledge-Flags

Pseudo-Code of the Search-Process

```
1.  Peer 0-User specifies 'information need' in a Query
2.  Peer 0 processes Query on local IR-Engine
 IF (local match was successful)
 → RANK and DISPLAY Hits → GOTO END.
 ELSE → GOTO 3.
3.  Peer 0 processes spatio-temporal range query on DSTI
 IF (ST-LOCATION == NULL)
 → use heuristic to calculate ST-MBR from own current location
 END IF
 request DSTI and receive a set of [0..n] knowledge-flags
 IF (n > 0) → GOTO 4.
 ELSE → heuristically calculate new ST-MBR → GOTO 3.
4.  Peer 0 selects [0..m] knowledge-flags from set where match WHAT == TRUE
 IF (m == 0)
 → select m knowledge-flags from set randomly
 ELSE IF (0 < m <= n)
 → forward query to the m peers associated with the knowledge-flags
 END IF
5.  Peers 1 receives Query
 process query refinement with own local knowledge (e.g. own GeoNER)
 process query on local IR-engine (see 2.) (also using own unpublished expert-links and
 information items
 rank set of n results (filtered by privacy heuristics)
 IF (n == 0) → GOTO 6.
 ELSE → GOTO 7.
6.  Peers 1 may inform own User about negative search result in local InfoSpace
 IF (intervention by User)
 → answer question with personal knowledge → n > 0
 ELSE → n = 0
7.  Peers 1 send a List with [0..n] records back to Peer 0
8.  Peer 0 receives the Lists from all Peers 1
 rank all answers from all lists of answers by consulting own expert-list.
 update own expert-list considering user feedback
```

Figure 5: Pseudo-code of the protocol for a first selection of appropriate peers (agents) with the help of the DSTI. Further iterations are subject to the respective agents.

where WHEN and WHERE may miss the T-RELATION and S-RELATION elements.

An answer in the system may be an indexed information-item plus a link for accessing the original information-item or the offer to establish a personal communication channel between the humans controlling the agents.

Partly summarizing and formalizing what has been previously discussed, figure 5 formally specifies the process of searching in our system.

This formal process contains several steps which e.g. have to employ suitable heuristics that need to be developed when developing the local IR systems. It is based on existing structures of Knowledge-Flags, indexing and extraction of spatio-temporal references of information-items and computation of Expertise-Cards through spatio-temporal clustering. For all these elements, current state of the art techniques exist that can be used.

5. EVALUATION

In order to evaluate key aspects of our approach, a prototype of the P2P protocol on the basis of the just discussed Quadtree as well as a simulation system for using the protocol was implemented. The main goal was to allow to evaluate and provide a proof of concept for our approach for selecting potentially relevant agents, allowing for a fine-tuning of the basic protocol. As Information-Items we used articles from Wikipedia. Dumps of German and English Wikipedia were read using the WikAPIdia API (Hecht 2009) and distributed to 2000 simulated agents in a partially disjoint way. For each agent, a geographic clustering on the set of his articles was conducted, thus creating his Expertise-Cards, which then were inserted into the spatial index. Using a small-world generator, we created a small world network of agents, which we used to create Expert-Links between the connected agents

Figure 6 shows a visualization of the DSTI as well as some spatial locations of agents and the Expert-Links between them, visualized as edges. (Regard that in our system the Expert-Links are spatio-temporally located as Expert-Link-Flags via the spatio-temporal reference of the location of the expertise of the target agent it points to. This is not shown in the figure.) The spatial locations of the agents are shown in the figure in the Quadtree. We see that the Quadtree is generally more dense in regions with more agents located in (and generally with more Expertise-Flags and Expert-Link-Flags in).

The simulation system allows for querying the distributed index from the point of view of an information seeking agent. The system shows that the P2P protocol based on the DSTI fulfills the requirements derived from our concept for a MAS. A qualitative evaluation with five evaluators showed that spatio-temporal reference as primary classification criterion is able to deliver relevant results for queries with and also without explicitly specified spatio-temporal reference. Precision and recall–based quantitative studies do not seem appropriate in this case, because key advantages of our approach, e.g. results that, due to small world effects, exhibit spatial, semantic or social relations to the query that have not explicitly been searched for, cannot be precisely captured by these measures. It is subject to future research to develop derived formal quality measures that can capture these advantages in an appropriate way. A meaningful quantitative evaluation will include a larger number of users and this developed quality measure.

Figure 6: Visualisation of Agents and Expert-Link edges in the DSTIs Quadtree

6. CONCLUSION

In this paper, we presented our approach to an alternative, socially and spatio-temporally sensitive IR, especially suitable for mobile environments as a reaction to the ever increasing interdependency between real and virtual world. The natural process of humans for information acquisition, their given location in space and time in connection with these processes as well as the cognitive understanding of these phenomena were an important basis for the essential considerations. In the first part of this contribution, we discussed problems and limitations of current centralized approaches to IR in view of an increasingly social and mobile Web. Secondly we discussed the principles of human IR and the relevance of spatio-temporal references of information and persons as a basis for a decentralized IR approach in our spirit. After that, we gave a short summary of recent approaches in Geo-IR and P2P-IR in order to classify and demarcate our approach. Another central aspect for federated socially sensitive IR are small world structures of the networks connecting the peers / agents and Tobler's first law (Tobler 1970) which were shortly reviewed. In a second part, we described our approach in more detail. We discussed spatio-temporal references as main classification criterion in our concept together with the central aspect that it implicitly conserves given semantic and social small world structures when embedding them into space and time. This gave rise to the notion of implicit social- and semantic spatio-temporal small worlds. We then discussed how the considerations discussed up to then are realized in architectural elements of our system. We presented a workflow of the developed P2P protocol used in retrieval processes in our system. Finally we discussed our qualitative evaluation.

Decentralized information retrieval in the spirit proposed here represents an interesting field of research. To our knowledge the incorporation of local heuristics and generally of specialized locally optimized IR systems based on P2P and the resulting possibilities for the improvement of quality, up-to-date-ness and personal quality of IR have not been regarded in great depth yet. The smooth and well performing interplay between the agents proposed here is also an interesting field of study for distributed systems research and multi-agent-research. Furthermore, declaratively specifying elements such as Expertise-Cards and interoperability for such IR concepts on a data-format level may be a relevant subject for the Semantic-Web community. Many other aspects, such as the precise structural properties of implicitly semantic, social and socially-semantically-combined spatio-temporal small worlds e.g. induced by such concepts as Expert-Links are promising fields of research which can be stimulated by federated social IR concepts in our spirit.

It is of special interest to generally investigate conflicting issues such as a massive P2P distribution and replication of content in view of increased performance versus keeping the data with your local agent and possibly not providing it when not online thus contributing to implementing a (partly) forgetting Web in contrast to a non-forgetting Web. We will also evaluate and further develop various heuristics with respect to the selection of suitable peers in case of missing or incomplete spatio-temporal elements of the query. It is also interesting to investigate the influence of information density and resulting tree-depth and respective quadrant-sizes of the respective distributed Quadtree on search strategies and clustering of information items as Expertises.

Elements of future work include research on suitably including models for measuring and using locality of information and more fine grained cognitive models of space into our approach. Another very important research direction is to investigate social access control mechanisms and privacy aspects of our model.

We wish to thank our students Henry Pötzl and Benjamin Koster for their valuable contributions. ◀

Literatur

Abell, J. C. (2009): Twitter Talks Data With Microsoft, Google - All Things Digital. (online: www.wired.com/epicenter/2009/10/twitter-talks-data-with-microsoft-google-all-things-digital/, Zugriff 10/2009).

Albaredes, G. (1992): A New Approach: User Oriented GIS. In: Proceedings of EGIS '92. Munich (1992): EGIS Foundation. ▷

Amitay, E.; Har'El, N.; Sivan, R.; Soffer, A. (2004). Web-a-where: geotagging web content. In: Proceedings of the 27th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Sheffield, United Kingdom, July 25-29, 2004). SIGIR '04. ACM, New York, pp. 273-280.

Andrade, L.; Silva, M. J. (2006): Indexing Structures for Geographic Web Retrieval. In: Proceedings of the Conference on Mobile and Ubiquitous Systems (CSMU'2006), Guimarães, Portugal.

Antoniou, G.; Van Harmelen, F. (2008): A Semantic Web Primer. MIT Press.

Bawa, M.; Bayardo, R. J. Jr.; Rajagopalan, S.; Shekita, E. (2003): Make it Fresh, Make it Quick: Searching a Network of Personal Webservers. Proc. WWW 2003, Budapest, Hungary.

Beckmann, N.; Kriegel, H.-P.; Schneider, R.; Seeger B.(1990): The R*-tree: An Efficient and Robust Access Method for Points and Rectangles. Proc. ACM SIGMOD'90, pp. 322-331.

Bliujut'e, R.; Jensen, C. S.; Saltenis, S.; Slivinskas, G. (1998): R-tree Based Indexing of Now-Relative Bitemporal Data. Proc. VLDB'98, pp. 345-356.

Brewington, B. E.; Cybenko, G. (2000): How dynamic is the Web? Proc. 9th International World Wide Web Conference, Amsterdam, Netherlands.

Brin, S.; Page, L. (1998): The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks 30(1-7), pp. 107-117.

Cho, J.; Garcia-Molina, H. (2002): Parallel crawlers. Proc. 11th International World Wide Web Conference, Honolulu, Hawaii, USA.

Dabek, F.; Brunskill, E.; Kaashoek, M.; Karger, D.; Morris, R.; Stoica, I.; Balakrishnan, H. (2001): Building Peer-to-Peer Systems With Chord, a Distributed Lookup Service. In: Proceedings of the 8th Workshop on Hot Topics in Operating Systems (HotOS-VIII) (2001).

Egenhofer, M. J.; Mark, D. M. (1995): Naïve Geography. In: Frank, A. U.; Kuhn, W.(Eds.) (1995): Spatial Information Theory: A theoretical foundation for GIS. Springer (LNCS 998), pp. 1-15.

Fabrikant, S. I.; Ruocco, M.; Middleton, R.; Montello, D. R.; Jörgensen, C. (2002). The First Law of Cognitive Geography: Distance and Similarity in Semantic Space. Proceedings, GIScience 2002, Boulder, CO, Sep. 25-28, 2002: pp. 31-33.

Fetterly, D.; Manasse, M.; Najork, M.; Wiener, J. (2003): A Large-Scale Study of the Evolution of Web Pages. Proc. 12th International World Wide Web Conference, Budapest, Hungary.

Gartner Inc. (2010): Gartner Highlights Key Predictions for IT Organizations and Users in 2010 and Beyond. (online: http://www.gartner.com/it/page.jsp?id=1278413, Zugriff 1/2010).

Gluck, M. (1995): Geospatial Information Needs of the General Public: Texts, Maps, and User's Tasks. Geographic Information Systems: Patrons, Maps and Spatial Information, pp 151-172.

Goodchild, M. (2004): The Validity and Usefulness of Laws in Geographic Information Science and Geography. Annals of the Association of American Geographers 94, pp. 300-303.

Groh, G. (2005): Ad-Hoc-Group Information Management in Mobile Communities - Identification and Applications. PhD thesis, Technische Universität München, Germany, Fakultät für Informatik.

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

Gibs, J. (2009): The Evolution of Content Discovery: From Portals to Search to Social Media; Nielsen Media Analytics (online: http://en-us.nielsen.com/etc/medialib/nielsen_dotcom/en_us/documents/pdf/webinars.Par.95561.File.pdf, Zugriff 4/2010)

Guttman, A. (1984): R-trees: a dynamic index structure for spatial searching. Proceedings of the 1984 ACM SIGMOD international conference on Management of data, June 18-21, Boston, Massachusetts.

Haase, P.; Siebes, R.; van Harmelen, F. (2004): Peer selection in peer-to-peer networks with semantic topologies. In: Bouzeghoub, M. (Eds.) (2004): Proceedings of the international conference on semantics in a networked world (ICNSW'04), Springer (LNCS 3226), pp. 108-125.

He, B.; Patel, M.; Zhang, Z.; Chang, K. (2007): Accessing the Deep Web: A Survey. Communications of the ACM 50 (2), pp. 94-101.

Hecht, B.; Raubal, M. (2008a): GeoSR: Geographically explore semantic relations in world knowledge. In: Bernard, L.; Friis-Christensen, A.; Pundt, H. (Eds.) (2008): Eleventh AGILE International Conference on Geographic Information Science, Springer, pp. 95-114.

Hecht, B.; Schöning, J. (2009b): Mapping the Zeitgeist. In: Miller, H.; Cova, T. (Eds.) (2009): 5th International Conference on Geographic Information Science . (Extended Abstract). Park City, Utah. United States.

Hecht, B.; Moxley, E. (2009): Terabytes of Tobler: Evaluating the First Law in a Massive, Domain-Neutral Representation of World Knowledge. Proc. COSIT '09. pp. 88-105.

Hecht, B.; Gergle, D. (2010a): On The "Localness" of User-Generated Content. Proc. 2010 ACM Conference on Computer Supported Cooperative Work. Savannah, Georgia, USA.

Hecht, B.; Gergle, D. (2010b): The Tower of Babel Meets Web 2.0: User-Generated Content and its Applications in a Multilingual Context. Proc. 28th ACM Conference on Human Factors in Computing Systems, Atlanta, GA.

Hevner, A. R.; March, S. T.; Park, J.; Ram, S. (2004): Design Science in Information Systems Research. MIS Quarterly, 28(1), pp. 75-106.

Hill, L. (2000): Core elements of digital gazetteers: Placenames, categories and footprints. In: Borbinha, J.; Baker, T. (Eds.) (2000): Proc. 4th European Conference, ECDL 2000 Lisabon, Portugal, September 18-20, 2000, Springer (LNCS 1923), pp. 280-290.

Jones, C. B.; Purves, R. S. (2008): Geographical information retrieval. International Journal of Geographical Information Science, Vol. 22, No. 3, pp. 219-228.

Joseph, S. (2002): NeuroGrid: Semantically Routing Queries in Peer-to-Peer Networks. In: Gregori, E.; Cherkasova, L.; Cugola,G.; Panzieri, F.; Picco, G. (Eds.) (2002): Web Engineering and Peer-to-Peer Computing Proc. NETWORKING 2002 Workshops Pisa, Italy, May 19-24, 2002, Revised Papers, Springer (LNCS 2376), pp. 202-214.

Kleinberg, J. (1998): Authoritative sources in a hyperlinked environment. In: Proc. Ninth annual ACM-SIAM symposium on Discrete algorithms. San Francisco, USA, pp: 668-677.

Kleinberg, J. (2000a): The small-world phenomenon: An algorithmic perspective. In: Proc. Annual ACM Symposium on Theory of Computing, Portland, Oregon, United States. pp. 163-170.

Kleinberg, J. (2000b): Navigation in a small world. Nature 406. pp. 845.

Kleinberg, J. (2004): The small-world phenomenon and decentralized search. A short essay as part of Math Awareness Month 2004, appearing in SIAM News 37(3), April 2004.

Larson, R. R.; Smith, L. C.; Gluck, M. (1996): Geographic Information Retrieval and Spatial Browsing. Geographic Information Systems: Patrons, Maps and Spatial Information, pp 81-123.

Larson, R. R.; Frontiera, P. (2004): Ranking and representation for geographic information retrieval. In: Purves R. and Jones C. B. (Eds) (2004): Proc. Workshop on Geographic Information Retrieval, SIGIR 2004 (Sheffield University, Sheffield).

Lawrence, S.; Giles, C. (1999): Accessibility of information on the Web. Nature 400, pp. 107-109.

Li, M.; Lee, W. C.; Sivasubramaniam, A. (2004): Semantic Small World: An Overlay Network for Peer-to-Peer Search. In: Proc. of the International Conference on Network Protocols (ICNP), pp. 228-238.

Liben-Nowell, D.; Novak, J.; Kumar, R.; Raghavon, P.; Tomkins, A. (2005): Geographic routing in social networks. PNAS 102(33), pp. 11623-11628.

Liberman, M.; Lin, J. (2009): You are where you edit: Locating Wikipedia contributors through edit histories. Proceedings of ICWSM'09, San Jose, USA.

Lua, E. K.; Crowcroft, J.; Pias, M. (2005): A Survey and Comparison of Peer-to-Peer Overlay Network Schemes. IEEE Communications Surveys and Tutorials 7(2).

Löser, A.; Tempich, C.; Quilitz, B.; Staab, S.; Balke, W. T.; Nejdl, W. (2005): Searching dynamic communities with personal indexes. In: Gil, Y.; Motta, E.; Benjamins, V. R.; Musen, M. A. (Eds.) (2005): Proc. 4th International Semantic Web Conference, ISWC 2005, Springer (LNCS 3729), pp. 491- 505.

Manning, C. D.; Raghavan, P.; Schütze, H. (2008): Introduction to Information Retrieval, Cambridge University Press.

Milgram, S. (1967): The small world problem. Psychology today, 1(61).

Miller, H. J. (2004): Tobler's First Law and Spatial Analysis. Annals of the Association of American Geographers 94, pp. 284-289.

Mondal, A.; Yilifu; Kitsuregawa, M. (2004): P2PR-tree: An R-tree-based spatial index for peer-to-peer environments. In: Proceedings of the International Workshop on Peer-to-Peer Computing and Databases (held in conjunction with EDBT), Heraklion-Crete, Greece.

Montello, D. R.; Fabrikant, S. I.; Ruocco, M.; Middleton, R. S. (2003): Testing the first law of cognitive geography on point-display spatializations. In: Kuhn, W.; Worboys, M. F.; Timpf, S. (Eds.): COSIT 2003. Springer (LNCS 2825), pp. 316-331.

Mountain, D.; MacFarlane, A. (2007): Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. Journal of Information Science, Vol. 33, Nr. 5, pp. 515-530.

Nahemov, G.; Lawton, M. (1975): Similarity and Propinquity in Friendship Formation. Journal of Personality and Social psychology 31(2), pp. 205-213.

Newman, M.; Barabasi, A.; Watts, D. J. (2006): The structure and dynamics of networks. Princeton University Press.

Ntoulas, A.; Cho, J.; Olston, C. (2004): What's new on the Web?: The evolution of the Web from a search engine perspective. Proc. 13th International World Wide Web Conference, New York, NY, USA.

Purves, R. S.; Clough, P. D. (2006): Judging spatial relevance and document location for geographic information retrieval. Proc. 4th International GIScience Conference, IfGI prints, Münster, Germany

Ratnasamy, S. et al. (2001): A Scalable Content Addressable Network. Proc. of the ACM SIGCOMM, pp. 161-172.

Risson, J.; Moors, T. (2006): Survey of Research towards Robust Peer-toPeer Netwrks: Search Methods. In: Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol. 50, Issue 17, pp. 3485-3521.

Rowstron, A.; Druschel, P. (2001): Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In: IFIP/ACM International Conference on Distributed Systems Platforms (Middleware), Springer (LNCS 2218), pp. 329-350.

Saltenis, S.; Jensen, C. S. (1999): R-Tree Based Indexing of General Spatio-Temporal Data, TR-45, TIMECENTER Technical Report (online: http://timecenter.cs.aau.dk/TimeCenterPublications/TR-45.pdf, Zugriff 2/2010).

Samet, H. (1984): The Quadtree and Related Hierachical Data Structures, Ausgabe 16. ACM, New York.

Sanderson, M.; Kohler, J. (2004): Analyzing geographic queries. Proc. ACM SIGIR Workshop on Geographic Information Retrieval, Sheffield, UK.

▷

Schmitz, C. (2005): Self-Organization of a Small World by Topic. Proc. of 1st International Workshop on Peer-to-Peer Knowledge Management.

Steinmetz, R.; Wehrle, K. (Eds.) (2004): Peer-to-Peer Systems and Applications, Springer (LNCS 3485).

Sui, D. Z. (2004): Tobler's First Law of Geography: A Big Idea for a Small World. Annals of the Association of American Geographers 94, pp. 269-277.

Tang, C.; Xu, Z.; Dwarkadas, S. (2003): Peer-to-peer information retrieval using self-organizing semantic overlay networks. In: Proceedings of ACM SIGCOMM, pp. 175-186.

Tanin, E.; Harwood, A.; Samet, H. (2007): Using a distributed quadtree index in peer-to-peer networks. VLDB Journal, Springer, 16(2), pp. 165-178.

Tobler, W. R. (1970): A computer movie simulating urban growth in the Detroit region. Economic Geography 46, pp. 234-240.

Tobler, W. R. (2004): On the First Law of Geography: A Reply. Annals of the Association of American Geographers 94, pp. 304-310.

Toffler, A. (1980): The Third Wave. The Classic Study of Tomorrow. Bantam Books, USA, New York.

Vaid, S.; Jones, C. B.; Joho, H.; Sanderson, M. (2005): Spatio-textual indexing for geographical search on the web. Proc. 9th International Symposium on Spatial and Temporal Databases, 22-24 August 2005, Angra dos Reis, Brazil, Springer, pp. 218-235.

Vestavik, Ø. (2004): Geographic Information Retrieval: An Overview. Dept. Computer and Information Science, Norwegian University of Technology and Science, Trondheim, Norwegian (online: http://www.idi.ntnu.no/~oyvindve/article.pdf, Zugriff 3/2010).

Watts, D. J.; Strogatz, S. H. (1998): Collective dynamics of 'small-world' networks. Nature 393 (6684), pp. 409-410.

Wellman, B.; Hogan, B. (2004): The immanent Internet. In: J. McKay (Ed.) (2004): Netting Citizens: Exploring Citizenship in a Digital Age. Edinburgh, St. Andrew Press, pp 55-80.

Wooldridge, M. (2001): Multi-Agent-Systems. Wiley & Sons.

Wu, L-S.; Akavipat, R.; Menczer, F.; Maguitman, A. G. (2007): Adaptive Peer-to-Peer social networks for distributed content based web search. In: Social Information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively. IDEA GROUP INC.

Yeung, C. A.; Liccardi, I.; Lu, K.; Seneviratne, O.; Berners-Lee, T. (2009): Decentralization: The Future of Online Social Networking. Proc. W3C Workshop on the Future of Social Networking, Barcelona, Spain.

Zhang, H.; Croft, W. B.; Levine, B.; Lesser, V. (2004): A multi-agent approach for peer-to-peer information retrieval. In: Proceedings of Third International Joint Conference on Autonomous Agents and Multi-Agent Systems.

GEOFORUM MV 2010 –VERNETZTE GEODATEN: VOM SENSOR ZUM WEB

von Lutz Kreßner

Der Verein für Geoinformationswirtschaft Mecklenburg-Vorpommern (GeoMV) veranstaltete am 26. und 27. April im Technologiepark Rostock/ Warnemünde das 6. GeoForum MV. Die zweitägige Konferenz, welche sich in diesem Jahr unter das Motto „Vernetzte Geodaten: vom Sensor zum Web" stellte, bringt einmal jährlich Vertreter aus Wissenschaft, Wirtschaft, Politik, Lehre und Forschung zusammen, um hier aktuelle technisch-wissenschaftliche Entwicklungen aus dem Umfeld der Geoinformationswirtschaft zu präsentieren.

Die Kernthemen der diesjährigen Veranstaltung erstreckten sich über eine große fachliche Bandbreite aus dem Bereich der Geo-Informationssysteme, wobei hier diesmal Fragestellungen aus den Schwerpunkten Geovisualisierung, 3D-Stadtmodelle, Open Street Map und vernetzte Geodaten dominierten. Hervorzuheben sind u.a. die Vorträge von Frau Prof. Dr. Dransch über die Visualisierung von Geodaten im raum-zeitlichen und attributiven Kontext der Geowissenschaften, die Vorstellung des Solardachkatasters Lage der Firma GEOPLEX GmbH (Gewinner des GeoBusiness Award 2009) durch Herrn Hilling sowie der Beitrag „Was ist eine gute Karte" von Herrn Prof. Dr.-Ing. Schiewe von der Hafencity Universität Hamburg, welcher sich mit den Herausforderungen bei der Qualitätsbeschreibung in der Geovisualisierung beschäftigte.

In einer parallel stattfindenden Firmenausstellung hatten regional, national und international agierende Unternehmen (u.a. GTA Geoinformatik GmbH, ESRI Deutschland GmbH) Gelegenheit ihre technischen Produkte und Dienstleistungen vorzustellen und auch zum gemeinsamen Erfahrungsaustausch anzuregen.

Zeitgleich zu den Fachvorträgen fanden am zweiten Konferenztag parallele Workshops statt, in denen sich die über 120 angemeldeten Teilnehmer unter anderem über bestehende Anforderungen und aktuelle Anwendungsbeispiele aber auch über Probleme sowie Potenziale aus den Themengebieten „Mobile Anwendungen", „Immobilien-Wirtschaft" sowie „Globale Umwelt- und Sicherheitsüberwachung" informieren konnten. Gastredner waren hier unter anderem Herr Schwarz vom DLR e.V. Neustrelitz („Ship Detection Service"), Herr Wiese (Northbit GmbH) mit dem Thema „NVE – Verwaltung und Vermarktung von Immobilien auf Basis von Landkarten" sowie Herr Dr. Grenzdörffer, welcher einen Einblick in den Stand und die Perspektiven von unmanned airborne vehicles (UAS) zur Geodatengewinnung gab.

Eine von vielen Teilnehmern besuchte gemeinsame Abendveranstaltung im Teepott Restaurant Warnemünde rundete die Konferenz ab und lud zum geselligen Treffen aber auch zur vertieften Diskussion ein.

Ein Tagungsband mit allen Beiträgen liegt bereits vor und kann direkt über den GITO Verlag (http://www.gito.de/) bestellt werden. Darüber hinaus stehen die einzelnen Vorträge auch über die Internetpräsenz der Veranstalter (www.geomv.de/geoforum/2010/beitraege.php) als PDF-Datei zur Verfügung. ◀