A Survey of Volunteered Open Geo-Knowledge Bases in the Semantic Web

Andrea Ballatore, David C. Wilson and Michela Bertolotto

Abstract Over the past decade, rapid advances in web technologies, coupled with innovative models of spatial data collection and consumption, have generated a robust growth in geo-referenced information, resulting in spatial information overload. Increasing 'geographic intelligence' in traditional text-based information retrieval has become a prominent approach to respond to this issue and to fulfill users' spatial information needs. Numerous efforts in the Semantic Geospatial Web, Volunteered Geographic Information (VGI), and the Linking Open Data initiative have converged in a constellation of open knowledge bases, freely available online. In this article, we survey these open knowledge bases, focusing on their geospatial dimension. Particular attention is devoted to the crucial issue of the quality of geo-knowledge bases, as well as of crowdsourced data. A new knowledge base, the OpenStreetMap Semantic Network, is outlined as our contribution to this area. Research directions in information integration and Geographic Information Retrieval (GIR) are then reviewed, with a critical discussion of their current limitations and future prospects.

Keywords: linked data; geographic information; geo-knowledge bases; data quality; Semantic Geospatial Web

Andrea Ballatore
School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland, e-mail: andrea.ballatore@ucd.ie

David C. Wilson
Department of Software and Information Systems, University of North Carolina, 9201 University City Boulevard, Charlotte, NC 28223-0001, USA, e-mail: davils@uncc.edu

Michela Bertolotto
School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland, e-mail: michela.bertolotto@ucd.ie

1 Introduction

In 1998, U.S. Vice President Al Gore delivered a speech at the California Science Center about what he named Digital Earth, a "multi-resolution, three-dimensional representation of the planet, into which we can embed vast quantities of geo-referenced data" [47, p. 89]. Much of the unprecedented amount of information produced and released on the Internet is about a specific place on the Earth. However, Gore pointed out, most of this informational wealth is generated and left untapped. Among the key aspects that would enable a more efficient exploitation of geo-data, interoperability and metadata were considered of particular importance. Multiple data sources should be combined together using a common framework, and metadata should describe online resources in a clear, standardised way [47].

Over the past 14 years, several geospatial initiatives have been undertaken, oriented towards the implementation of the Digital Earth [45]. The Open Geospatial Consortium (OGC)[1] has defined and promoted several standards to distribute geographic data, while the Global Spatial Data Infrastructure (GSDI) association aims at fostering "spatial data infrastructures that support sustainable social, economic, and environmental systems integrated from local to global scales" [1, p. 1]. Despite these efforts, standard formats are often ignored in favour of application-specific formats. As Fonseca et al. put it, heterogeneity emerges spontaneously in a free market of ideas and products, and standards cannot reduce it by decree [33].

The spectacular growth of unstructured information online has affected all domains, prompting Tim Berners-Lee to envisage the advent of the so-called Semantic Web [10]. The Semantic Web project aims to develop a standard semantic format to describe online data, originating a network of machine-readable, semantically clear documents. Data semantics is expressed in predicate logic-based languages such as RDF,[2] in large collections of statements about real world entities. This vision was further formalised through the Linked Data initiative, which promotes the release of datasets in an inter-connected web of semantic data [12].

The information explosion in geographic data has not only been quantitative, but also qualitative [113]. With the rise of Web 2.0, Internet users have become active producers of geo-referenced information, utilising collaborative web tools in large projects [85]. Several collaborative efforts emerged to create and maintain large datasets, resulting in crowdsourcing, impacting initially on non-spatial information and subsequently also on the geographic domain [61]. In the geospatial context, the term 'neogeography' has been used in order to refer to this rapid and complex nexus of technological and social practices [106]. Goodchild termed the crowdsourcing of geographic information as Volunteered Geographic Information (VGI), emphasising its production through voluntary labour [44]. Haklay et al. have surveyed VGI projects [53], while Coleman et al. have discussed the practices and motivations of 'produsers', users/producers of geographic data [25]. In addition, Sui used the term

---

[1] http://www.opengeospatial.org (acc. June 5, 2012)

[2] http://www.w3.org/RDF (acc. June 5, 2012)

'wikification' to describe the practice of crowdsourcing of non-textual data, emulating the Wikipedia model in the geographic domain [104].

The impact of neogeography is not restricted to non-profit, academic organisations. Private institutions such as Google, Microsoft, and Yahoo! are progressively offering facilities for sharing geo-data, expanding their services beyond the routing systems that dominated the first phase of web-based Geographic Information Systems (GIS).[3] In this sense, geo-wikification is identifiable in the growth of web services allowing users, with some degree of freedom, to create or edit spatial data. As Priedhorsky notes, however, most interactive geo-services are essentially 'digital graffiti,' i.e. annotations on a static geographic image [90]. Beyond the specificities of each case, it can be argued that all neogeographic and VGI phenomena share the characteristics of being volunteered, crowdsourced, wikified, and web-based.

Even though the popular claim that 80% of information is geo-referenced has been questioned [51], it can be stated safely that, over the past decade, geo-information has experienced a remarkable growth [71]. As happened in other fields subject to an information explosion and subsequently to information overload, the issue of semantics of geo-data – or lack thereof – has become critical. The deluge of semantically ambiguous geo-data caused Egenhofer to advocate the emergence of a Semantic Geospatial Web, a spatial extension of the Semantic Web [29]. In Egenhofer's view, this new framework for geospatial information retrieval should rely on the semantics of spatial and terminological ontologies. Thanks to inter-operable semantic representations of the data, the Semantic Geospatial Web will increase the relevance and quality of results in geographic retrieval systems.

As a result of the synergy between crowdsourcing, VGI, and the Semantic Geospatial Web, several large-scale collaborative projects have emerged. While Wikipedia[4] is without doubt the most visible text-based crowdsourcing project, OpenStreetMap (OSM) has applied the wiki model to create an open world vector map [54]. Several geo-knowledge bases have then been created by structuring existing datasets into Semantic Web formats: the projects LinkedGeoData, GeoNames, and GeoWordNet are salient examples [6, 41]. Research efforts have been undertaken on the development, maintenance, and merging of open geo-knowledge bases, to enhance the geographic intelligence of information retrieval systems, beyond the traditional text-based techniques [33, 113, 6, 31].

Moreover, GIR has attempted to increase the geographic awareness of text-based information retrieval systems. On top of traditional flat gazetteers (dictionaries of toponyms and geo-coordinates), GIR has started exploiting geo-knowledge bases to reduce the ambiguity of geographic terms and enable spatial reasoning [67, 86]. Despite these efforts, the knowledge contained in such computational artifacts is left largely untapped. We believe that these open geo-knowledge bases have potential in addressing the challenges of GIR, and deserve particular attention. For this reason, we provide a survey of currently active knowledge bases with particular emphasis on

---

[3] See http://maps.google.com, http://www.bing.com/maps, http://maps.yahoo.com

[4] http://www.wikipedia.org (acc. June 5, 2012)

their geospatial content, and we review the state of the art in information integration and GIR, including our contribution to these areas.

The remainder of this chapter is organised as follows: Section 2 surveys the constellation of online open knowledge bases containing geographic knowledge. Applications of geo-knowledge bases are discussed in Section 3, from recent efforts in ontology alignment and merging (Section 3.1), to ontology-powered GIR (Section 3.2). Section 4 presents our work in this area, describing the OSM Semantic Network[5] and the semantic expansion of the OSM dataset, connecting it to DBpedia, in order to enrich spatial data with a richer knowledge base [8]. The issue of quality of geo-knowledge bases is discussed in Section 5, and Section 6 offers a review of current limitations of these computational artifacts that need particular consideration in their usage. Finally, Section 7 discusses the challenges lying ahead in this field and the further research required to identify solutions to these challenges.

2 Survey of Open Linked Geo-Knowledge Bases

This section provides a survey of open, collaborative geo-knowledge bases, which constitute an important part of semantic technologies. To avoid terminological confusion, it is beneficial to provide a definition of the related and sometimes overlapping terms used in knowledge representation. A 'knowledge base' is a collection of facts about a domain of interest, typically organised to perform automatic inferences [50]. A knowledge base contains a terminological conceptualisation (typically called 'ontology') and a set of individuals. Widely used both in philosophy and in computer science, the meaning of the term 'ontology' is particularly difficult to define [99]. Among the many definitions, "an explicit specification of a conceptualization" and "shared understanding of some domain of interest" are of particular relevance, as they stress the presence of an explicit formalisation, and the general aim of being understood within a given domain [50, p. 587]. Winter notes that ontologies became part of Geographic Information Science (GIScience) towards the end of the 20th century [112].

A 'thesaurus' is a list of words grouped together according to similarity of their meaning [92], whilst a digital 'gazetteer' is specifically geographic, and contains toponyms, categories, and spatial footprints of geographic features [59]. In the Web 2.0 jargon, a 'folksonomy' is a crowdsourced classification of online objects, based on an open tagging process [109]. Finally, a 'semantic network,' a term which originated in psychology, is a graph whose vertices represent concepts, and whose edges represent semantic relations between concepts [91].[6] We define a 'geo-knowledge base' as a knowledge base containing some geographic information.

---

[5] http://wiki.openstreetmap.org/wiki/OSMSemanticNetwork (acc. June 5, 2012)

[6] Unlike ontologies, semantic networks focus on psycho-linguistic aspects of the terms. However, some knowledge bases, such as WordNet, defy this distinction by showing aspects of both ontologies and semantic networks.

In the context of geographic information, a knowledge base is generally made up of an ontology, defining classes and their relationships (abstract geographic concepts such as 'lake'), and then populated with instances of these classes, generally referring to individual entities (e.g. Lake Victoria and Lake Balaton). In this survey, we restrict the scope to projects having global coverage, discussing their spatial content. These knowledge bases are the result of combined efforts in crowdsourcing, VGI, and the Semantic Geospatial Web, and offer useful resources for GIR, and other areas of geo-science.

The Semantic Web and the Linked Data initiatives promote the adoption of semantic formats, which can be used to add an open, machine readable semantic structure to online data [12, 46]. In this context, several collaborative projects have emerged, resulting in a growing number of freely available geo-knowledge bases. Among these numerous resources, we focus on eleven datasets that have a global scope (as opposed to local projects), are mostly generated through crowdsourcing, released under Creative Commons/Open Database licences,[7] and which are available as fully downloadable dumps in popular Semantic Web formats such as OWL and RDF. Some of the selected projects are focused specifically on geographic data (e.g. GeoNames and OpenStreetMap), while others are more general-purpose but contain valuable geographic knowledge (e.g. DBpedia and Freebase). These knowledge bases provide open datasets, and are inter-connected with one another. Our own contribution to this area of research, the OSM Semantic Network, is described in Section 4. Relevant characteristics of each knowledge base are summarised in Table 1.

CONCEPTNET   This semantic network is focused on natural language processing and understanding [58]. ConceptNet is a large semantic network, whose nodes represent concepts in the form of words or short phrases of natural language. The graph edges represent labelled relationships. Each statement in ConceptNet has justifications pointing to it, explaining where it comes from and how reliable the information seems to be. The ontology includes 1.6 million assertions gathered from Wikipedia, Wiktionary, WordNet, and the 700,000 sentences from the Open Mind Common Sense project [97]. Efforts to encode ConceptNet in RDF are being undertaken [48].

DBPEDIA   One of the leading projects of the Semantic Web, DBpedia is a Semantic Web version of Wikipedia [5]. The knowledge base currently contains 3.6 million entities, encoded in a billion RDF triples, including 526,000 places. As DBpedia is strongly interconnected with other knowledge bases (e.g. WordNet W3C, GeoNames, LinkedGeoData), it is considered the central hub of Linked Data.

FREEBASE   Designed as an open repository of structured data, Freebase allows web communities to build data-driven applications [13]. The knowledge base is structured around terms (classes), and unique entities (instances), where an entity can be a specific person, a place, or a thing, and is described by facts. It currently contains 22 million entities, of which 1 million are locations. As entities are

---

[7] See http://creativecommons.org and http://opendatacommons.org (acc. June 5, 2012)

Project Name	Year*	Type & Content	Data sources	Formats
CONCEPTNET	2000	Ontology, semantic network; 1.6 million assertions, 700,000 natural language sentences	Wikipedia, WordNet, and others	JSON
DBPEDIA	2007	Ontology, semantic network; 320 classes, 740K Wikipedia types, 3.6M entities, 1 billion triples	Wikipedia	OWL/RDF
FREEBASE	2007	Ontology, knowledge base; 22M+ entities, 1M locations	Crowdsourced	Tab separated text
GEONAMES	2006	Gazetteer; 650 classes, 10M+ toponyms	Gazetteers, Wikipedia, crowdsourced	OWL/RDF
GEOWORDNET	2010	Semantic network, thesaurus, gazetteer; 330 classes, 3.6M entities	WordNet, GeoNames, MultiWordNet	RDF
LINKEDGEODATA	2009	Gazetteer; 1K classes, 380M geographic entities	OpenStreetMap	RDF
OPENCYC	1984	Ontology, semantic network; 50K classes, 300K facts	Expert-authored Cyc knowledge base	OWL/RDF
OPENSTREETMAP	2004	Vector map, gazetteer; User-defined tags, 1.2B nodes, 114M ways	Crowdsourced, free GIS datasets	XML
WIKIPEDIA	2001	Semantic network, dictionary, thesaurus; Semi-structured (infoboxes), 3.9M articles in English	Crowdsourced	XML
WORDNET	1985	Semantic network, dictionary, thesaurus; 117K synsets	Expert-authored knowledge base	OWL/RDF
YAGO	2006	Ontology, semantic network; 10M+ entities, 460M facts	Wikipedia, GeoNames, WordNet	RDF

Table 1 A survey of open ontologies. All of these projects are currently active, release open data, have global scope, and are interconnected with other projects. *Beginning of the project.

described by facts corresponding to a directed graph, it can be easily converted into RDF.

GEONAMES   Combining multiple data sources, GeoNames aims at offering a large, volunteered gazetteer.[8] The knowledge base contains over 10 million toponyms, structured in 650 classes. GeoNames integrates geographical data such as names of places in various languages, elevation, and population. The data is collected from traditional gazetteers such as National Geospatial-Intelligence Agency's (NGA) and the U.S. Geological Survey Geographic Names Information System (GNIS), and crowdsourced online.

GEOWORDNET   GeoWordNet is the result of the integration of WordNet, GeoNames and the Italian part of MultiWordNet [41]. It is a hybrid project, combining a semantic network, a dictionary, a thesaurus, and a gazetteer. It was developed in response to the limited WordNet coverage of geospatial information and lack of concept grounding with spatial coordinates. The knowledge base contains 3.6 million entities, 9.1 million relations between entities, 334 geographic concepts, and 13,000 (English and Italian) alternative entity names, for a total of 53 million RDF triples.

LINKEDGEODATA (LGD)   Since OpenStreetMap has gathered a large collection of geographic data, LinkedGeoData is an effort to republish it in the Semantic Web context [6]. The OSM vector dataset is expressed in RDF according to the Linked Open Data principles, resulting in a large spatial knowledge base. The knowledge base currently contains 350 million nodes, 30 million ways (polygons and polylines in the OSM terminology), resulting in 2 billion RDF triples. Some entities are linked with the corresponding ones in DBpedia.

OPENCYC   This is the open source version of Cyc, a long running artificial intelligence project, aimed at providing a general knowledge base and common sense reasoning engine.[9] Even though OpenCyc covers a limited number of geographic instances, it contains a rich representation of specialised geographic classes, such as *salt lake* and *monsoon forest*. The OpenCyc classes are interlinked with DBpedia nodes and Wikipedia articles.

OPENSTREETMAP (OSM)   The OSM project aims at constructing a world vector map [54]. The leading VGI initiative, the dataset represents the entire planet, gathering data from existing datasets, GPS traces, and crowdsourced knowledge. To date, the vector dataset contains 1.2 billion nodes (points), and 115 million ways (polygons and polylines).

WIKIPEDIA   A collaborative writing project, Wikipedia is a multilingual, universal encyclopedia, and has become the most visible crowdsourcing phenomenon.[10] The English version currently contains 3.9 million articles, resulting in a 2 billion-word corpus. Because of high connectivity between its articles, Wikipedia is sometimes used as a semantic network [102]. This vast repository of general knowledge has been used for different purposes, including semantic similarity

---

[8] http://www.geonames.org (acc. June 5, 2012)

[9] http://www.cyc.com/opencyc, http://sw.opencyc.org (acc. June 5, 2012)

[10] http://www.wikipedia.org (acc. June 5, 2012)

and ontology extraction [110, 84]. The project has also attracted interest in the area of GIScience [3].

WORDNET   Initially conceived as a lexical database for machine translation, WordNet has become a widely used resource in various branches of computer science, where it is used as a semantic network and as an ontology [32]. Currently it contains 117,000 'synsets', groups of synonyms corresponding to a concept, connected to other concepts through several semantic relations. The dataset has been encoded and released in RDF, becoming a highly linked knowledge base in the web of Linked Open Data.[11] Even though the spatial content of WordNet is limited, the ontology holds a high quality, expert-authored conceptualisation of geographic concepts.

YAGO   Yet Another Great Ontology (YAGO) is a large knowledge base extracted from Wikipedia and Wordnet [103]. Recently YAGO has been extended with data from GeoNames, with particular emphasis on the spatial and temporal dimensions [60]. The current version of the knowledge base contains 10 million entities, encoded in 460 million facts. YAGO is inter-linked with DBpedia and Freebase.

Figure 1 presents the constellation of geo-knowledge bases, showing a schematised data path from the data producers to the knowledge bases. Bearing in mind the complexity of these collaborative processes, the main actors in this constellation, involved in the production of information and the generation of open linked knowledge bases, can be grouped as follows:

1. **Data providers.** Traditionally, geographic data was collected exclusively by experts and professionals in large public and private institutions. As Web 2.0 and VGI have emerged, a new category of non-expert users/producer ('produsers') has entered the production process [25]. Crowdsourced primary sources include contributions from a wide variety of information producers, ranging from experts operating within public and private institutions to non-expert, unpaid, pro-active users.

2. **Primary sources.** Projects such as Wikipedia and OSM collect a large amount of information about the world through crowdsourced efforts. On the other hand, primary sources such as WordNet are expert-authored, while other projects combine both crowdsourcing and expert control. Most knowledge bases rely heavily on these primary sources, often aligning and merging them into larger knowledge bases. Inconsistencies and contradictions in primary sources can be propagated onto the derived knowledge bases. For example, an incorrect piece of information in a Wikipedia article will be also found in DBpedia and YAGO. For this reason, assessing the quality of these primary sources bears particular importance (see Section 5).

3. **Geo-knowledge bases.** Typically, open knowledge bases consist of structured and aggregated versions of existing semi-structured or unstructured primary sources. However, some datasets lie at the boundary between primary sources

---

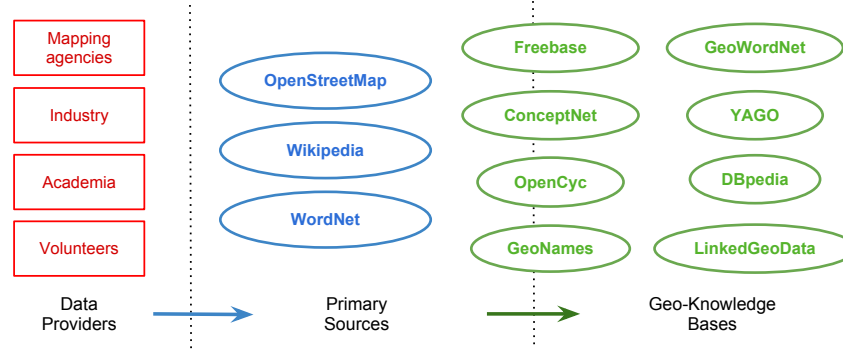[11] http://www.w3.org/TR/wordnet-rdf (acc. June 5, 2012)

Fig. 1 The constellation of open geo-knowledge bases. The data path is schematised from the data providers to semi-structured primary sources, and finally structured into knowledge bases. Some projects defy classification by producing new knowledge in structured knowledge bases, and extracting knowledge from primary sources.

and knowledge bases, as they are both interlinked with existing knowledge bases and produce new data through crowdsourcing and expert contributions (e.g. Freebase and OpenCyc). Several knowledge bases encode the same primary data into different formalisms, such as DBpedia and YAGO.

These three actors are part of an open system, in which more or less structured data flows in complex patterns that determine the nature, quality and limitations of the resulting projects. Investigations on such collaborative open processes have been carried out, both in the area of general crowdsourcing and VGI [25]. The next section covers applications where these knowledge bases play an important role, in particular in relation to ontology alignment, and GIR.

3 Open Geo-Knowledge Bases in action

Since the late 1990s, geospatial knowledge bases have been of fundamental importance in many geographic applications [112], including semantic geographic information systems [2], GIR [3], and toponym disambiguation [87]. In general, geo-knowledge bases are used to achieve semantic interoperability between local geographic datasets modelled on incompatible ontologies. Geo-knowledge bases can also be useful in cases where advanced geographic knowledge is necessary to interpret unclear, fuzzy spatial information queries and needs [57, 41]. We focus initially on ontology alignment and merging (Section 3.1), and we subsequently discuss the usage of ontologies in GIR (Section 3.2). Our own contributions to this area of research are presented in Section 4.

3.1 Mapping, aligning, and merging Geo-Knowledge Bases

Online geo-data is stored in many different formats, leading to a radical heterogeneity of data formats, ontologies, and semantic models [33]. Standards by the Open Geospatial Consortium (OGC) and International Organization for Standardization (ISO) are used in geospatial modelling, while other standards are developed and promoted by the World Wide Web Consortium (W3C) in the context of web technologies. To date, despite some initial efforts, there is no clear sign of convergence towards broad adoption of joint standards between those two communities [95]. From an information retrieval perspective, the issue of coverage is critical, as GISs want to access geo-data from as many sources as possible in a consistent way. For these reasons, the field of integration of geo-knowledge bases has received a lot of attention, and is currently an active research area [73, 88].

The integration of heterogenous data sources relies on the semantic matching of geo-knowledge bases, which present similar conceptualisations of geographic entities using different syntax, structure, and semantics [107]. According to Giunchiglia et al., considering ontologies as graphs, their alignment consists of the production of a "set of correspondences between the nodes of the graphs that correspond semantically to each other" [42, p. 1]. Combining geo-knowledge bases poses unresolved challenges, mostly due to the dynamic nature of online open datasets, semantic ambiguity, and inconsistent use of the same vocabulary, which is defined by Vaccari et al. as the 'semantic heterogeneity problem' [107].

Such an integration can operate at the class level (e.g. identify and inter-link the concept 'lake' in all the data sources), or at the instance level (e.g. find the entities representing Lake Ontario in all of the ontologies). Moreover, the semantic mapping can be applied to data, such as geo-ontologies, vector datasets, and gazetteers, or to services, in particular web services, focusing on the public interface signatures [33, 66]. Choi et al., surveying this area of research, have defined three categories of ontology mapping. The first category includes mapping between local ontologies and a higher-level ontology. The second category of mapping is performed between local ontologies. The third category of mapping is part of ontology merging, in which existing ontologies are combined in a bigger ontology [22].

In order to identify semantically close classes and instances in different ontologies, semantic similarity measures are particularly useful. Semantic similarity measures specific for geographic classes have been surveyed by Schwering [96]. She classifies the existing measures into geometric, feature, network, alignment, and transformational models. Janowicz et al. have proposed a formal framework for geo-semantic similarity [65]. This new framework responds to the ambiguity and lack of clear theoretical grounds that characterise the area of semantic similarity measurement.

Some form of ontology alignment and merging, either partly or fully automatic, has been utilised to generate most of the geo-knowledge bases surveyed in Section 2. The process leading to the creation GeoWordNet, for example, relies on the alignment between GeoNames and WordNet at the class level. Some of the concepts modelled by GeoNames were not defined in WordNet, prompting the creation

of new synsets. After the ontologies were aligned at the class level, it was possible to align them at the instance level, resulting in the new, integrated ontology [41].

Similarly, LinkedGeoData has mapped some of its instances to corresponding entities in DBpedia, by aligning the ontologies along feature type, spatial distance, and name similarity [6]. The ontology YAGO is assembled by aligning WordNet synsets with the less structured Wikipedia articles [103]. Along similar lines, Buscaldi et al. have linked existing gazetteers with WordNet, and Wikipedia [21]. Their system extracts place names from the freely available Geonet Names Server (GNS) and the Geographic Names Information System (GNIS). Relevant place names are then filtered and enriched using semantic knowledge in Wikipedia and WordNet. The particular challenge that this system addresses is the combination of semantically flat place names with nodes in complex semantic networks.

The system GeoMergeP uses a layered architecture to combine and merge local ontologies, through the ISO 19100 standard [17]. Surveying recent ontology integration works, Bucella et al. identify three main techniques, which at times are used in isolation, and at times in combination: (1) top-level ontology, (2) logical inferences, and (3) matching/similarity functions. GeoMergeP combines all of the three approaches to overcome their limitations.

While most approaches are top-down, the integration can be a bottom-up process. A bottom-up ontology alignment, focused on geographic linked data, has been carried out by Parundekar et al. [88]. This work adopts the approach of common extension comparison, i.e. two classes in different ontologies are considered similar if they are linked by similar instances, to align DBpedia, GeoNames, and LinkedGeoData (see Section 2). The mapping is done through an alignment hypothesis, built bottom-up, starting from instance pairs, up to the most general classes in the ontology.

Ontology alignment is also used by Smart et al. to combine multiple gazetteers through a common, high-level ontology [98]. Their Geo-Feature Integration module combines toponyms from OSM, GeoNames, Wikipedia, and other sources into a unified gazetteer. The module relies on spatial and textual similarity to match places across the selected data sources. In addition to traditional text similarity measures, this system uses the SoundEx algorithm to match phonetically similar sounding terms to detect alternative – and wrong – place name spellings.

Once the geo-knowledge bases have been integrated, they can be used to support various spatial tasks. In particular, GIR has emerged as a prominent area that can benefit from geo-knowledge bases [68]. The next Section surveys recent work in the area of geo-knowledge bases applied to GIR.

3.2 Ontology-powered Geographic Information Retrieval (GIR)

Information retrieval (IR) is a vast and rapidly evolving area of computer science. Manning et al. define it as "finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections"

[78, p. 1]. Users' information needs often include some spatial information, such as a geo-location, a street name, and so on. In the context of ever-growing online information, the geographic dimension of information has become a promising way to increase the chances of meeting information needs.

Traditionally, search engines have treated spatial-related terms as any other textual information. Over the past decade, however, the area of GIR has emerged to develop techniques to give geographic information a special treatment, increasing the system's geographic intelligence [67]. Geographic information is often implicit in the documents: broad geographic entities are omitted when they are assumed to be known to the readers, e.g. Ireland is not mentioned when referring to Dublin in the Irish media. Toponyms (place names) also have a high degree of semantic ambiguity, as there are many terms to indicate the same geographical entity in different cultural and social groups, and several places have the same name (e.g. more than 40 North-American towns are called Greenville). Moreover, toponyms pose particular challenges across different natural languages, where historical and spelling variations are very common [86].

Geo-knowledge bases have been identified as a promising support tool to develop more sophisticated GIR systems [68]. While describing their project YAGO, Weikum et al. advocate the usage of knowledge bases to go beyond the limitations of current keyword-based search engines [111, 103]. As they put it, the main challenge is "how to extract the important facts from the Web and organize them into an explicit knowledge base that captures entities and semantic relationships among them" [111, p. 61]. To illustrate YAGO's knowledge representation, the entity representing Max Planck is displayed, including geographic knowledge about the town where the German physicist was born. The underlying intuition is that geographic knowledge is generally not provided explicitly, therefore knowledge bases can be used to discover implicit connections between entities.

Ontologies have been used in information retrieval to increase the system intelligence. In GIR, Lutz and Klien described an ontology-based system [74]. A shared vocabulary is used to translate queries across multiple ontologies, without defining a full global ontology. Their GIR system allows for user-friendly queries, translating generic queries to specific, local geo-knowledge bases. This is accomplished in a transparent way using Description Logics (DL), a family of knowledge representation languages based on first-order logic that has gained popularity in Semantic Web applications [7]. Fouad et al. have devised a location-based service to retrieve semantic information based on the user's location [34]. Their application performs keyword-based queries on DBpedia, LinkedGeoData, and GeoNames, with the aim of displaying semantically enriched web maps. Furthermore, in the area of location-based services, DBpedia Mobile demonstrates the possibility of obtaining rich semantic information about the user's surroundings [9].

Among others, we regard the following areas to be particularly promising as application domains for geo-knowledge bases:

Named Entity Recognition and Classification (NERC). Several systems rely on NERC techniques to identify location, people and organisations names in raw text. Nadeau and Sekine surveyed NERC techniques, from the field inception in

1991 to 2006, and discussed the main strategies to evaluate them [83]. While most NERC approaches are at least partly supervised, Cimiano and Völker have developed an ontology-based, unsupervised NERC procedure [23]. Overell has developed an approach to recognise and classify geo-referenced entities in Wikipedia articles [86].

Toponym disambiguation.   As many geographic locations share the same name, resolving the correct referent in a given context is far from being trivial. Toponym disambiguation is a specific case of proper name disambiguation, where the proper names refer to explicit or implicit spatial relationships. Knowledge-based techniques exploit geo-knowledge bases [20]. For example, Overell and Rüger have utilised Wikipedia as a knowledge base to perform place name disambiguation [87]. A co-occurrence model is extracted from Wikipedia to provide not only a list of synonyms for each location, but also the context in which each synonym is used. Toponym disambiguation is tightly connected to the issue of toponym resolution.

Toponym resolution.   By definition, a toponym refers to a geographic location. For this purpose geo-coding (toponyms to locations) and reverse geo-coding (locations to toponyms) services have been built using open geo-knowledge bases, extending traditional gazetteers with richer semantic structures. Smart et al., for example, have integrated several toponym sources into an ontological geo-coding and reverse geo-coding service [98]. In order to associate geographic locations to entities, Odon et al. have extracted textual evidence from Wikipedia articles [3]. Using Wikipedia as a semantic network, the importance of the entities is assessed by their connectivity with other entities. In this way, a representation of the geographic content of Wikipedia articles can be obtained. A related task is that of 'spatial co-reference resolution,' i.e. determining whether two digital representations refer to the same real-world entity. De Tré et al. developed approaches to detect co-referent features based on possibility theory, and applied it to the issue of duplicate detection [16, 26].

Spatial footprints.   Text or multimedia documents can be associated with a spatial footprint, which can be a simple geo-coordinate, a minimum bounding rectangle, or a complex polygon. Suitable spatial footprints can be computed and indexed, allowing for efficient retrieval and combination with pure text-based indexing. Fu et al. have devised an ontology and footprint-based query expansion mechanism [36]. Spatial entities are identified in a geographic ontology, and the spatial footprint of terms is computed and used in the retrieval process. Similarly, Vaid et al. have described different indexing approaches for text documents, showing that spatial indexing can enrich pure textual indexing to search large collections of text documents [108]. In the same area, Martins et al. discuss the 'geo-scope' of a text document, which is essentially a spatial footprint to be matched against the query footprint [79].

Spatial reasoning.   In information retrieval, queries are often expressed in natural language. Words such as 'in' and 'near' can convey important spatial semantics, which should be taken into account to meet users' information needs. GIR systems can utilise ontologies to carry out inferences on such geospatial

hints. For example, the fuzzy query 'lakes near Dublin' could be translated into a 'within' spatial query with a radius appropriate to the user context. Fu et al. have conducted work in this area in the frame of the project SPIRIT (Spatially-Aware Information Retrieval on the Internet) [36]. SPIRIT relies on dedicated ontologies to interpret spatial relationships between query terms, in the format ⟨*what*,*relationship*,*where*⟩. Valid relationships include, among others, 'near', 'north', and 'outside of'. More recently, the user context emerged as an important element that should be handled by GIR systems.

User context.   Any information need, whether containing a geographic dimension or not, is relative to a user context. For example, a user might want to retrieve all the Italian restaurants in London to conduct a socio-economic analysis, or simply to go out for dinner. The user context contains diverse information about the user, such as interests, current location, habits, language, computational device, etc, and can be exploited to refine semantic similarity measures and GIR [69]. Keßler et al. have devised a semantic language to enrich OWL with context sensitivity [70]. The *Geo-Finder* system extracts fuzzy spatial footprints from text documents, determining the scope of the search based on the user location and speed [14]. The spatial context is taken into account by Mobile Geotumba, a GIR system optimised for handheld devices, to retrieve local information [35].

All of the aforementioned areas of research are active, and open geo-knowledge bases reviewed in Section 2 can provide useful tools to explore novel approaches. Despite the promising results obtained in several works, such ontologies have clear limitations with respect to quality that should not be ignored by researchers and general users alike. The quality and limitations of geo-knowledge bases will be discussed in Sections 5 and 6.

The next section describes our contribution to the area of geo-knowledge bases, in particular presenting the OSM Semantic Network and a semantic enhancement technique to link OSM entities to DBpedia.

4 The OSM Semantic Network

In this section we describe our own contribution to the area of geo-knowledge bases and information integration. The OSM Semantic Network is a resource that we have extracted from OpenStreetMap (OSM) data to provide a semantic support tool. OSM is the leading project of VGI, and its vector dataset has been discussed, evaluated, and utilised in various contexts [54, 52]. The OSM Semantic Network is extracted through a dedicated web crawler we have developed, and provides a detailed representation of the conceptualisation underlying OSM.

In OSM, the semantics of map entities is described through tags, fragments of text with a key and a value (e.g. *amenity=park*, *name='Central Park'*). Such tags are proposed, defined, and discussed on a wiki website, which hosts detailed definitions

RDF Prefix	Vertex Type	Instances
`osmwiki:Key:`	OSM Key.	884
`osmwiki:Tag:`	OSM Tag.	2,047
`osmwiki:Proposed_features`	OSM Proposed Tag.	340
other	LGD and Wikipedia nodes.*	1,398

RDF Prefix	Edge Type	Instances
`osmwiki:link`	Internal link.	11,982
`osmwiki:valueLabel`	A value of a OSM tag.	2,926
`osmwiki:keyLabel`	OSM key.	2,251
`rdf:rdf-schema#comment`	OSM Tag description.	1,892
`osmwiki:key`	Link to OSM key page.	1,891
`osmwiki:combinedWith`	Tag is combined with target tag.	1,257
`osmwiki:link`	A link to a Wikipedia page.	1,118
`osmwiki:redirect`	Redirect to a OSM wiki page.	478
`osmwiki:implies`	Tag implies target tag.	97

Table 2 The OSM Semantic Network (extracted on June 10, 2011). Vertices marked with * are leaf vertices, i.e. have only incoming edges. 'osmwiki:' stands for http://wiki.openstreetmap.org/wiki/

and usage guidelines for the project.[12] In the wiki pages, users often link the OSM tags to similar concepts in Wikipedia. Overall, the tagging process is deliberately informal and open to revision. Contributors are encouraged to stick to well-known tags, but the creation of new tags is not discouraged, resulting in highly dynamic – and often inconsistent – semantics.

In the context of the Linked Open Data, LinkedGeoData (LGD) has converted and published the OSM vector dataset in RDF, linking it to a formally defined ontology [6]. However, the LinkedGeoData ontology is a simple, shallow tree representing tags. To the best of our knowledge, the rich semantic information on the OSM wiki website has not been included. In order to fill this knowledge gap, we have developed an open source tool, the OSM Wiki Crawler, which extracts an RDF graph from the OSM Wiki website. The crawler extracts a semantic network in RDF, whose vertices represent tags, and edges relationships between tags. Tags are linked to Wikipedia pages, and to existing LinkedGeoData classes. The edge labels specify a number of different relationships between vertices, ranging from a generic internal link (`link`) to a logical implication (`implies`). The detailed content of the current RDF graph is summarised in Table 2. In addition to the OSM Wiki Crawler, pre-extracted RDF graphs are available online.[13] Among other applications, this ontology can be used as a support to compute semantic similarity between tags [65], as well as aligning OSM and LinkedGeoData to other geo-knowledge bases [41].

In the context of ontology alignment, we have developed an integration technique between LinkedGeoData and DBpedia, matching geographic features across

---

[12] http://wiki.openstreetmap.org/wiki/Map_Features (acc. June 5, 2012)

[13] http://wiki.openstreetmap.org/wiki/OSMSemanticNetwork (acc. June 5, 2012)

the datasets. As discussed in Section 2, some LinkedGeoData instances are linked to corresponding nodes in DBpedia, in particular cities, airports, lakes, and other well-defined entities. This alignment was performed in the context of pessimistic assumptions, favouring precision over coverage. As a result, only a small subset of OSM objects is linked to DBpedia. Thus, to obtain a wider coverage, we have adopted more flexible heuristics, based on geographic proximity and a tag matching mechanism based on key words. A web application was built to allow users to visually explore the OSM dataset, and extract DBpedia nodes and concepts related to the geographic entities displayed in the current web map. A preliminary evaluation, published in [8], suggests a promising performance of this ontology-based system, but further work is needed to explore its strengths and weaknesses.

5 The quality of crowdsourced Geo-Knowledge Bases

In order to utilise open knowledge bases successfully, it is crucial to assess their quality with respect to the user's requirements. For example, a geo-knowledge base might have sufficient quality to enrich the semantics of a web-based GIR system, but is likely to fail to meet the standards needed by the transport industry. Assessing the quality of knowledge bases can benefit project owners, contributors, and users, indicating criteria to select the best available resource for a given task and highlighting limitations and design flaws.

A crucial trade-off in the geo-knowledge bases discussed in this survey is between coverage and precision. Wikipedia-based ontologies such as DBpedia and YAGO cannot aim at pristine perfection, but can still obtain a reasonable precision [103]. On the other hand, expert-authored resources such as WordNet have very high precision, but are unable to compete with the coverage of crowdsourced projects. A similar trade-off applies to the geospatial dimension: geo-knowledge bases can either reach high, expert-validated spatial quality, or can be updated very frequently by a large number of volunteers, but it is difficult for these two elements to co-exist.

In recent years, several quantitative approaches to assess the quality of an ontology have been discussed [43, 15, 101, 37, 100, 49]. In our view, the approaches to evaluate the quality of geo-knowledge bases can be classified in four families:

1. **Manual evaluation**: domain experts and intended users analyse manually the knowledge base, highlighting issues and giving qualitative judgements on the mapping between the knowledge base and the real world domain that the knowledge base is supposed to capture [40]. Although human subjects can easily detect design flaws in the schema, the labour cost of human experts can make it impractical. Moreover, even in the presence of considerable resources, large knowledge bases cannot be fully evaluated manually, and automatic methods are needed.
2. **Within-knowledge-base evaluation**: particular properties of a knowledge base are observed without comparison with external sources. These approaches are based on relationship patterns, distributional patterns, and logical inconsistencies [19, 105]. Although this approach is inexpensive, and can be adopted for

any knowledge base, its effectiveness is largely context-dependent. The average connectivity between objects, for instance, can vary across different domains, without being a reliable indicator of quality.

3. **Between-knowledge-base evaluation**: two knowledge bases covering the same domain can be compared, cross-checking their quality. If one of the two knowledge bases has comparatively high quality, it can be used as a 'ground truth' or 'gold standard' to validate the other. For example, datasets collected and validate by national mapping agencies tend to obtain higher spatial accuracy than equivalent crowdsourced data [52]. Clearly, this approach cannot be used when comparable knowledge bases are unavailable, a rather common situation.

4. **Application-based evaluation**: ultimately, knowledge bases are designed and populated to provide support for real-world applications. Hence, an approach consists of observing the performance of a task with and without a given knowledge base, and measuring the differential as an indicator of quality. In this framework, different knowledge bases can be compared indirectly, bearing in mind that knowledge bases can obtain varying performances on different tasks. Strasunskas and Tomassen have proposed a scheme to evaluate the 'ontology fitness' with respect to search tasks [101].

In practice, quality assessment strategies can combine these four approaches in different ways, along multiple dimensions. The quality of a knowledge base can be measured at the class level and at the instance level, looking at the statistical properties of the knowledge base. For example, it is possible to have a solid, well-designed schema but noisy, insufficient instances, or vice-versa. A combination of these two aspects can offer a comprehensive picture of the ontology quality.

Specific approaches to ontology evaluation focus on a set of dimensions. Tartir et al. [105], for example, outlined a within-ontology approach based on a triangular model, in which three dimensions of quality can be observed: between the real world and the schema, between the real world and the knowledge base, and between the schema and knowledge base. In their formulation, metrics for schema quality include 'relationship richness,' 'attribute richness,' and 'inheritance richness,' while instance metrics capture 'class importance,' 'cohesion,' 'connectivity,' and 'readability.' Logical inconsistencies in the knowledge base can also be detected and used to measure quality [4]. For example, a knowledge base can contain the conflicting statements 'Canada *southOf* USA' and 'USA *southOf* Canada.'

Moreover, Burton-Jones et al. addressed the issue of ontology quality from a semiotic viewpoint, proposing a within-ontology evaluation framework [19]. The quality is observed from four perspectives: 'syntactic quality' (richness of lexicon and correctness), 'semantic quality' (interpretability, consistency and clarity), 'pragmatic quality' (comprehensiveness, accuracy and relevance), and 'social quality' (authority and history). An overall indicator of quality is obtained with a linear combination of these four dimensions.

In the context of the open geo-knowledge bases that we have described in Section 2, the quality of primary sources such as Wikipedia and OSM has a great impact of the derived ontologies. The reliability of Wikipedia has fostered a major academic and intellectual debate, without reaching a monolithic verdict [75]. A typical way

of assessing the quality of Wikipedia is based on a between-knowledge-base comparison of a random sample of articles against a well-established, expert-authored encyclopedia [40]. The results indicate that Wikipedia has excellent coverage, but the quality of its articles can vary from poor to excellent. Hu et al. have proposed within-ontology quality measures for Wikipedia articles, based on the authoritativeness of the contributors [62].

Although false information, hoaxes and spam are generally corrected in a timely manner, Wikipedia articles at a given time can always have errors being introduced and removed. Therefore, when a snapshot of the Wikipedia website is stored and analysed, any particular article might happen to be captured right after being vandalised or after a thorough revision by a domain expert. To date, no easy solution to this issue exists.

Assessing the quality of geographic data is a well known area of GIScience, traditionally developed in the framework of cartography [18, 63]. Several dimensions can be observed to assess the quality of spatial information, including positional accuracy (how accurate the object location is with respect to the real world), completeness (how many objects are represented in the map versus all the existing objects), and logical consistency (duplicate objects, inconsistent topological relationships, etc.). Moreover, semantic aspects are particularly important for GIR, such as attribute and semantic accuracy, which focus on the quality of the metadata. The temporal quality, i.e. the rate and accuracy of updates, bears particular importance in several geospatial applications [52].

Indeed, the advent of VGI introduces additional challenges. The quality and reliability of OSM has been debated since its inception, and is now considered a critical research area for VGI [11, 82]. Like other crowdsourced projects, OSM has experienced recurring and extensive vandalism, urging the project founder to call for action [24]. Allegations have been put forward that some vandalism might be carried out by corporate competitors [38]. This sort of 'spatial vandalism' in open data poses peculiar challenges for project administrators, and has not yet been studied on a systematic basis.

Analogously to Wikipedia, precision and coverage of the OSM spatial data can vary greatly. An approach to quantify quality consists of adopting a map from a trusted source (e.g. a national mapping agency), and comparing it with OSM. Thus, Haklay have compared a sample from the OSM vector dataset against the corresponding data from the British Ordnance Survey [52]. OSM obtains a positional accuracy of 70%, with drops to 20%, a range that Haklay considers to be "not dissimilar to commercial datasets" [52, p. 700]. Along similar lines, Mooney et al. have conducted a quality analysis on a European subset of OSM [82]. Their study confirms the high variability in the data quality, identifying several geographical divides: rural and low-income areas tend to have lower coverage than wealthy, urban areas; natural features tend to be less covered than man-made features.

To date, the lack of standardised 'fitness' metrics to indicate the quality of open geo-knowledge bases makes their adoption problematic, particularly in areas in which the requirements are strict, e.g. logistics and transport. However, mainly for economic reasons, a number of online services are moving from commercial Web

maps to VGI data sources – the popular social network FourSquare being the most prominent case – indicating a rising trust in crowdsourced geographic data [56].

6 Current limitations of Geo-Knowledge Bases and GIR

Given the promise of geo-knowledge bases, GIR and the Semantic Web in general, it is important to be aware of the current limitations and drawbacks of such technologies. The Semantic Web is a broad and ambitious project that has made undeniable progress, but many of its issues are largely unresolved [81]. Polleres et al. have identified critical problems affecting the web of Linked Open Data, ranging from cases where there is too little data, poor quality data, or too much data [89]. We identify issues affecting the usability of linked geo-knowledge bases, restricting the discussion to aspects relevant to the ontologies reviewed in Section 2.

Ambiguity. Because of the wide variety of data linked by ontologies, the same vocabulary can have very different usages depending on the context. A paradigmatic case is the `owl:sameAs` predicate, which has become ambiguous in real datasets [55]. The difficulties in specifying geographic information share a common root in the complexity of the concept of place in natural languages. The conceptualisation of place is a cultural and language-dependent process, is intrinsically vague, refers to ever-shifting cultural borders, depends on other complex concepts, and is influenced by the context of usage [94]. Moreover, the web of open data lacks a meta-ontology framework to describe ontologies in a unified way [64].

Coverage. In some cases there is too little data, and missing entities or links prevent queries from retrieving results. When an entity has not been published in RDF and loaded in a public triple repository, it is simply unreachable. RDF adoption online is sparse, and most RDF triples are coming from mass imports from unstructured or semi-structured datasets [89]. When using open ontologies, the coverage/quality dilemma has to be taken into account: increasing coverage normally entails a drop in quality, and vice-versa. Projects aiming at global coverage often stumble upon the difficulty of keeping large knowledge bases in the same coherent semantic framework. Coverage also varies depending on fine-grained, project-specific aspects. In OSM, for example, man-made features are generally better covered than natural features [82]. The coverage of the interlinking between ontologies, can also show high variability, leaving vast areas of ontologies unlinked [89].

Quality. Most geo-knowledge bases contain a vast amount of data imported from crowdsourced projects. As discussed in Section 5, while crowdsourcing has clear advantages in terms of coverage and cost, precision is inevitably neglected. Moreover, when inconsistent, incomplete or inaccurate information is entered in Wikipedia or OSM, it will be propagated into DBpedia, YAGO, LinkedGeoData, and many other derived ontologies. The quality of VGI and crowdsourced data in general is hotly debated, and high variability has to be expected (see Section 5).

The difficulties related to creating, maintaining and interpreting metadata were bluntly but persuasively described in the 'Metacrap' article by Doctorow [27]. However, several ontology quality metrics have been devised. Stransunkas and Tomassen, while presenting a framework for ontology evaluation for information retrieval, survey existing ontology metrics [101]. Beside formal quality metrics based on structure, coherence, and other aspects, an open geo-ontology can be evaluated indirectly on the basis of results obtained in real-world tasks.

Expressivity. Modelling geographic knowledge into an ontology poses specific challenges. RDF triplification, however simple it might be in most cases, can be very complex and counter-intuitive for certain facts [89]. This issue is due to well-known representational limitations of semantic networks. Additionally, OWL expressivity for spatial data is very limited. As Abdelmoty et al. pointed out, OWL does not support spatial types, and common spatial operations such as distance are not available [2]. For spatial reasoning, OWL has to be used in conjunction with spatial databases, preventing a seamless integration with existing infrastructures [28].

Complexity. Spatial reasoning has been often identified as a fundamental instrument to increase intelligence in GIS [30]. However, applying complex spatial reasoning over large geo-knowledge bases poses remarkable challenges. Even in an ideal situation – data without noise and logical contradictions – reasoning in OWL Full is undecidable, and OWL DL is not designed to reason over massive, distributed datasets [89]. Further research is needed to enable efficient spatial reasoning in noisy, large, distributed knowledge bases.

Several efforts are being undertaken to tackle these issues in the context of the Linked Data initiatives [89]. However, it is reasonable to assume that the presence of noise, varying quality, and limited expressivity can be reduced but never fully resolved. Therefore, when developing applications relying on open geo-knowledge bases, caution is needed in order to deal with unexpected contradictions, inconsistencies, ambiguity, and a varying amount of noise in the data.

A prominent application area for geo-knowledge bases, GIR is a relatively young discipline, and its achievements are particularly difficult to assess [80]. Most of the works in the area present a preliminary evaluation, leaving the effectiveness of the approaches to be verified empirically in real world applications. To date, the most important large-scale evaluation is represented by the four GeoCLEF challenges, run from 2005 to 2008 [39, 77]. Focus was placed on open data in GikiCLEF 2009, an evaluation contest conceived to explore cultural and linguistic issues in Wikipedia-based GIR [93].

The driving intuition behind such initiatives is that adding geographic knowledge to an IR system would improve its performance when dealing with information needs with a spatial component. However, as Mandl noted, complex GIR systems have not consistently obtained better results than geographically naive systems [76]. According to Leveling, the contradictory results of GeoCLEF show possible areas of research that might improve the overall results of GIR, strengthening the usage of natural language processing with semantic indexing, handling metonyms, and topological relations beyond simple inclusion [72].

7 Conclusions and future work

In this chapter we presented a survey on recent advances in open geo-knowledge bases and GIR. In Section 1, we framed these areas of research in the combined visions about the Digital Earth, the Semantic Geospatial Web, and the emergence of Volunteered Geographic Information, which have changed the face of geographic information over the past decade [29, 44]. The linked open geo-resources available online that we discuss in this chapter are realising, at least in part, the vision of the 'collaboratory,' a collaborative geo-laboratory envisaged by Al Gore in 1998 to promote the development of geographic digital technologies [47].

A survey of free, open geo-knowledge bases with global coverage was presented in Section 2, including GeoNames, DBpedia, YAGO, GeoWordNet, ConceptNet, and others. Those knowledge bases are created by extracting knowledge from Wikipedia, OSM and traditional GIS data sources, merging different knowledge bases. Section 3.1 provides an overview of recent work in the area of geo-ontology alignment and merging. In order to cope with the growing amount of online geographic information, GIR has emerged. Section 3.2 surveys recent work in the usage of geo-knowledge bases to increase the geographic intelligence of GIR systems. Our own contribution to the area of open geo-knowledge bases, the OSM Semantic Network and an OSM/DBpedia alignment approach, is subsequently outlined in Section 4.

Section 5 surveyed the existing strategies to assess the quality of geo-knowledge bases, with particular emphasis on the quality of crowdsourced data sources. Despite undeniable advances towards the Semantic Geospatial Web and the increased coverage and quality of open geo-knowledge bases, it is important to recognise its current limitations. Section 6 highlights current issues which researchers using open geo-knowledge bases frequently encounter, identifying the core issues in coverage, quality, expressivity, and complexity of geo-knowledge bases. Similarly, current GIR systems have not met the expected increase in performance over traditional information retrieval, indicating that geographic intelligence needs refinement to become effective in its applications [76].

These issues notwithstanding, promising applications of open geo-knowledge bases are to be found in GIR, ontology alignment, toponym resolution, and related areas. In this respect, it can be argued that the most effective way to counter scepticism lies not only in formal, academic evaluations such as GeoCLEF, but in the production and dissemination of usable web applications for Internet users. For this purpose, more collaboration with the human computer interaction community might help devise appropriate interfaces to interact with open geo-data, exploiting these knowledge bases in convincing ways [76]. Work on open geo-knowledge bases should never lose contact with the ultimate stakeholders in information retrieval systems, the human users with their diversified and often unexpected information needs.

References

1. Strategic Plan 2009–2013. Tech. rep., Global Spatial Data Infrastructure Association (2009). URL http://portal.gsdi.org/files/?artifact_id=544
2. Abdelmoty, A., Smart, P., Jones, C.: Building place ontologies for the semantic web: issues and approaches. In: Proceedings of the 4th ACM workshop on Geographical Information Retrieval, pp. 7–12. ACM (2007)
3. Odon de Alencar, R., Davis Jr, C., Gonçalves, M.: Geographical classification of documents using evidence from Wikipedia. In: Proceedings of the 6th ACM Workshop on Geographic Information Retrieval, GIR'10, pp. 12:1–12:8. ACM (2010)
4. Arpinar, I., Giriloganathan, K., Aleman-Meza, B.: Ontology quality by detection of conflicts in metadata. In: Proceedings of the 4th International Workshop on Evaluation of Ontologies for the Web, EON 2006, pp. 1–7. IW3C2 (2006)
5. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A nucleus for a web of open data. In: The Semantic Web, *LNCS*, vol. 4825, pp. 722–735. Springer (2007)
6. Auer, S., Lehmann, J., Hellmann, S.: LinkedGeoData: Adding a Spatial Dimension to the Web of Data. In: Proceedings of the International Semantic Web Conference, ISWC 09, *LNCS*, vol. 5823, pp. 731–746. Springer (2009)
7. Baader, F., Horrocks, I., Sattler, U.: Description Logics as Ontology Languages for the Semantic Web. In: Mechanizing Mathematical Reasoning, *LNCS*, vol. 2605, pp. 228–248. Springer (2005)
8. Ballatore, A., Bertolotto, M.: Semantically Enriching VGI in Support of Implicit Feedback Analysis. In: Proceedings of the Web and Wireless Geographical Information Systems International Symposium (W2GIS 2011), *LNCS*, vol. 6574, pp. 78–93. Springer (2011)
9. Becker, C., Bizer, C.: DBpedia mobile: A location-enabled linked data browser. In: Proceedings of the WWW2008 Workshop on Linked Data on the Web, LDOW2008, *CEUR Workshop Proceedings*, vol. 369 (2008)
10. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American **284**(5), 28–37 (2001)
11. Bishr, M., Kuhn, W.: Geospatial Information Bottom-Up: A Matter of Trust and Semantics. In: The European Information Society: Leading the Way with Geo-information, Proceedings of the 10th AGILE Conference, LNGC, pp. 365–387. Springer (2007)
12. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story So Far. International Journal on Semantic Web and Information Systems. **5**(3), 1–22 (2009)
13. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, pp. 1247–1250. ACM (2008)
14. Bordogna, G., Ghisalberti, G., Psaila, G.: Geographic information retrieval: Modeling uncertainty of user's context. Fuzzy Sets and Systems **196**, 105–124 (2011)
15. Brank, J., Grobelnik, M., Mladenić, D.: A survey of ontology evaluation techniques. In: Proceedings of the Conference on Data Mining and Data Warehouses, SiKDD 2005, pp. 166–169. Information Society (2005)
16. Bronselaer, A., De Tré, G.: Semantical evaluators. In: Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, pp. 663–668. EUSFLAT (2009)

17. Buccella, A., Cechich, A., Gendarmi, D., Lanubile, F., Semeraro, G., Colagrossi, A.: GeoMergeP: Geographic information integration through enriched ontology matching. New Generation Computing **28**(1), 41–71 (2010)

18. Burrough, P., Rijn, R., Rikken, M.: Spatial Data Quality and Error Analysis: GIS Functions and Environmental Modeling. In: GIS and Environmental Modelling: Progress and Research Issues, pp. 29–34. John Wiley & Sons, Hoboken, NJ (1996)

19. Burton-Jones, A., Storey, V., Sugumaran, V., Ahluwalia, P.: A semiotic metrics suite for assessing the quality of ontologies. Data & Knowledge Engineering **55**(1), 84–102 (2005)

20. Buscaldi, D.: Approaches to disambiguating toponyms. SIGSPATIAL Special **3**(2), 16–19 (2011)

21. Buscaldi, D., Rosso, P., Peris, P.: Inferring Geographical Ontologies from Multiple Resources for Geographical Information Retrieval. In: 3rd Workshop on Geographic Information Retrieval, SIGIR. ACM (2006)

22. Choi, N., Song, I., Han, H.: A survey on ontology mapping. ACM Sigmod Record **35**(3), 34–41 (2006)

23. Cimiano, P., Völker, J.: Towards large-scale, open-domain and ontology-based named entity classification. In: Recent Advances in Natural Language Processing, RANLP 2005, pp. 166–172. ACL (2005)

24. Coast, S.: Enough is enough: disinfecting OSM from poisonous people. OpenGeoData (August 10, 2010). URL http://opengeodata.org/enough-is-enough-disinfecting-osm-from-poison

25. Coleman, D., Georgiadou, Y., Labonte, J.: Volunteered Geographic Information: the nature and motivation of produsers. International Journal of Spatial Data Infrastructures Research **4**(2009), 332–358 (2009)

26. De Tré, G., Bronselaer, A., Matthé, T., Van de Weghe, N., De Maeyer, P.: Consistently Handling Geographical User Data: Context-Dependent Detection of Co-located POIs. In: Information Processing and Management of Uncertainty in Knowledge-Based Systems: 13th International Conference, IPMU 2010, *LNCS*, vol. 6178, pp. 81–85. Springer (2010)

27. Doctorow, C.: Metacrap: Putting the torch to seven straw-men of the meta-utopia (2001). URL http://www.well.com/ doctorow/metacrap.htm

28. Dolbear, C., Hart, G., Goodwin, J.: What OWL has done for geography and why we don't need it to map read. In: Proceedings of the OWLED*06 Workshop on OWL: Experiences and Directions, *CEUR Workshop Proceedings*, vol. 216, pp. 10–11 (2006)

29. Egenhofer, M.: Toward the Semantic Geospatial Web. In: Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems, pp. 1–4. ACM (2002)

30. Egenhofer, M., Mark, D.: Naive geography. In: A.U. Frank, W. Kuhn (eds.) Spatial Information Theory: A Theoretical Basis for GIS, *LNCS*, vol. 1329, pp. 1–15. Springer (1995)

31. Euzenat, J., Ferrara, A., Meilicke, C., Nikolov, A., Pane, J., Scharffe, F., Shvaiko, P., et al.: Results of the Ontology Alignment Evaluation Initiative 2010. Tech. rep., Ontology Alignment Evaluation Initiative (2010)

32. Fellbaum, C. (ed.): WordNet: An electronic lexical database. MIT press, Cambridge, MA (1998)

33. Fonseca, F., Egenhofer, M., Agouris, P., Câmara, G.: Using ontologies for integrated geographic information systems. Transactions in GIS **6**(3), 231–257 (2002)

34. Fouad, R., Badr, N., Talha, H., Hashem, M.: On Location-Centric Semantic Information Retrieval in Ubiquitous Computing Environments. International Journal of Electrical & Computer Sciences **10**(6), 1–7 (2010)

35. Freitas, S., Afonso, A., Silva, M.: Mobile Geotumba: Geographic information retrieval system for mobile devices. In: Proceedings of the 4th MiNEMA workshop in Sintra, pp. 83–87. Department of Informatics, University of Lisbon (2006)

36. Fu, G., Jones, C., Abdelmoty, A.: Ontology-based spatial query expansion in information retrieval. In: On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, *LNCS*, vol. 3761, pp. 1466–1482. Springer (2005)

37. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: Modelling ontology evaluation and validation. The Semantic Web: Research and Applications **4011**, 140–154 (2006)
38. Garling, C.: Google Workers Caught 'Vandalizing' Open Source Maps. Wired (January 17, 2012). URL http://www.wired.com/wiredenterprise/2012/01/osm-google-accusation
39. Gey, F., Larson, R., Sanderson, M., Joho, H., Clough, P., Petras, V.: GeoCLEF: the CLEF 2005 cross-language geographic information retrieval track overview. In: Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evalution Forum, CLEF 2005, *LNCS*, vol. 4022, pp. 908–919. Springer (2006)
40. Giles, J.: Internet encyclopaedias go head to head. Nature **438**(7070), 900–901 (2005)
41. Giunchiglia, F., Maltese, V., Farazi, F., Dutta, B.: GeoWordNet: A Resource for Geo-Spatial Applications. In: The Semantic Web: Research and Applications, ESWC 2010, *LNCS*, vol. 6088, pp. 121–136. Springer (2010)
42. Giunchiglia, F., Yatskevich, M., Shvaiko, P.: Semantic matching: Algorithms and implementation. In: Journal on Data Semantics IX, *LNCS*, vol. 4601, pp. 1–38. Springer (2007)
43. Gómez-Pérez, A.: Evaluation of ontologies. International Journal of intelligent systems **16**(3), 391–409 (2001)
44. Goodchild, M.: Citizens as Sensors: the world of volunteered geography. GeoJournal **69**(4), 211–221 (2007)
45. Goodchild, M.: NeoGeography and the nature of geographic expertise. Journal of Location Based Services **3**(2), 82–96 (2009)
46. Goodwin, J., Dolbear, C., Hart, G.: Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. Transactions in GIS **12**, 19–30 (2008)
47. Gore, A.: The Digital Earth: Understanding our planet in the 21st century. The Australian Surveyor **43**(2), 89–91 (1998)
48. Grassi, M., Piazza, F.: Towards an RDF encoding of ConceptNet. In: Advances in Neural Networks: 8th International Symposium on Neural Networks, ISNN 2011, *LNCS*, vol. 6675, pp. 558–565. Springer (2011)
49. Guarino, N.: Toward a formal evaluation of ontology quality. IEEE intelligent Systems **19**(4), 78–79 (2004)
50. Guarino, N., Giaretta, P.: Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: N. Mars (ed.) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing, pp. 25–32. IOS Press, Amsterdam, Netherlands (1995)
51. Hahmann, S., Burghardt, D., Weber, B.: "80% of All Information is Geospatially Referenced"??? Towards a Research Framework: Using the Semantic Web for (In)Validating this Famous Geo Assertion. In: Proceedings of the 14th AGILE International Conference on Geographic Information Science, 158, pp. 1–9. Association Geographic Information Laboratories Europe (2011)
52. Haklay, M.: How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and Planning B: Planning and Design **37**(4), 682–703 (2010)
53. Haklay, M., Singleton, A., Parker, C.: Web Mapping 2.0: The Neogeography of the GeoWeb. Geography Compass **2**(6), 2011–2039 (2008)
54. Haklay, M., Weber, P.: OpenStreetMap: User-Generated Street Maps. IEEE Pervasive Computing **7**(4), 12–18 (2008)
55. Halpin, H., Hayes, P.: When owl:sameAs isn't the same: An analysis of identity links on the Semantic Web. In: Linked Data on the Web WWW2010 Workshop, LDOW2010, *CEUR Workshop Proceedings*, vol. 628, pp. 1–5 (2010)
56. Hardy, Q.: Facing Fees, Some Sites Are Bypassing Google Maps. New York Times (March 20, 2012). URL http://www.nytimes.com/2012/03/20/technology/many-sites-chart-a-new-course-as-google-expands-fees.html
57. Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., Riedemann, C.: Semantic interoperability: A central issue for sharing geographic information. The Annals of Regional Science **33**(2), 213–232 (1999)

58. Havasi, C., Speer, R., Alonso, J.: ConceptNet 3: A Flexible, Multilingual Semantic Network for Common Sense Knowledge. In: Recent Advances in Natural Language Processing, RANLP 2007, pp. 27–29. ACL (2007)

59. Hill, L.: Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. In: Research and Advanced Technology for Digital Libraries, *LNCS*, vol. 1923, pp. 280–290. Springer (2000)

60. Hoffart, J., Suchanek, F., Berberich, K., Lewis-Kelham, E., De Melo, G., Weikum, G.: YAGO2: exploring and querying world knowledge in time, space, context, and many languages. In: Proceedings of the 20th International Conference Companion on World Wide Web, pp. 229–232. ACM (2011)

61. Howe, J.: The Rise of Crowdsourcing. Wired magazine **14**(6), 1–4 (2006)

62. Hu, M., Lim, E., Sun, A., Lauw, H., Vuong, B.: Measuring Article Quality in Wikipedia: Models and Evaluation. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pp. 243–252. ACM (2007)

63. Hunter, G., Bregt, A., Heuvelink, G., Bruin, S., Virrantaus, K.: Spatial data quality: problems and prospects. In: Research Trends in Geographic Information Science, LNGC, pp. 101–121. Springer (2009)

64. Jain, P., Hitzler, P., Yeh, P., Verma, K., Sheth, A.: Linked data is merely more data. In: AAAI Spring Symposium on Linked Data Meets Artificial Intelligence, pp. 82–86. AAAI (2010)

65. Janowicz, K., Raubal, M., Kuhn, W.: The semantics of similarity in geographic information retrieval. Journal of Spatial Information Science **2**(1), 29–57 (2011)

66. Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maué, P., Stasch, C.: Semantic enablement for spatial data infrastructures. Transactions in GIS **14**(2), 111–129 (2010)

67. Jones, C., Alani, H., Tudhope, D.: Geographical Information Retrieval with Ontologies of Place. In: Spatial Information Theory, *LNCS*, vol. 2205, pp. 322–335. Springer (2001)

68. Jones, C., Purves, R.: GIR'05 2005 ACM workshop on Geographical Information Retrieval. In: ACM SIGIR Forum, vol. 40, pp. 34–37. ACM (2006)

69. Keßler, C.: Similarity measurement in context. In: Proceedings of the 6th International and Interdisciplinary Conference on Modeling and Using Context, *LNCS*, vol. 4635, pp. 277–290. Springer (2007)

70. Keßler, C., Raubal, M., Wosniok, C.: Semantic Rules for Context-Aware Geographical Information Retrieval. In: Smart Sensing and Context, *LNCS*, vol. 4793, pp. 77–92. Springer (2009)

71. Kitsuregawa, M., Nishida, T.: Special issue on information explosion. New Generation Computing **28**(3), 207–215 (2010)

72. Leveling, J.: Challenges for Indexing in GIR. SIGSPATIAL Special **3**(2), 29–32 (2011)

73. Lopez-Pellicer, F., Silva, M., Chaves, M., Javier Zarazaga-Soria, F., Muro-Medrano, P.: Geo Linked Data. In: Database and Expert Systems Applications, DEXA 2010, *LNCS*, vol. 6261, pp. 495–502. Springer (2011)

74. Lutz, M., Klien, E.: Ontology-based retrieval of geographic information. International Journal of Geographical Information Science **20**(3), 233–260 (2006)

75. Magnus, P.: On trusting Wikipedia. Episteme **6**(1), 74–90 (2009)

76. Mandl, T.: Evaluating GIR: geography-oriented or user-oriented? SIGSPATIAL Special **3**(2), 42–45 (2011)

77. Mandl, T., Gey, F., Di Nunzio, G., Ferro, N., Larson, R., Sanderson, M., Santos, D., Womser-Hacker, C., Xie, X.: GeoCLEF 2007: The CLEF 2007 cross-language geographic information retrieval track overview. In: Advances in Multilingual and Multimodal Information Retrieval, *LNCS*, vol. 5152, pp. 745–772. Springer (2008)

78. Manning, C., Raghavan, P., Schutze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge, England (2008)

79. Martins, B., Silva, M., Andrade, L.: Indexing and ranking in Geo-IR systems. In: Proceedings of the 2005 Workshop On Geographic Information Retrieval, GIR'05, pp. 31–34. ACM (2005)

80. Martins, B., Silva, M., Chaves, M.: Challenges and resources for evaluating geographical IR. In: Proceedings of the 2005 Workshop On Geographic Information Retrieval, GIR'05, pp. 65–69. ACM (2005)

81. Millard, I., Glaser, H., Salvadores, M., Shadbolt, N.: Consuming multiple linked data sources: Challenges and Experiences. In: First International Workshop on Consuming Linked Data, COLD2010, *CEUR Workshop Proceedings*, vol. 665, pp. 1–12 (2010)

82. Mooney, P., Corcoran, P., Winstanley, A.: Towards quality metrics for OpenStreetMap. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 514–517. ACM (2010)

83. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes **30**(1), 3–26 (2007)

84. Nakayama, K., Hara, T., Nishio, S.: Wikipedia Link Structure and Text Mining for Semantic Relation Extraction. Towards a Huge Scale Global Web Ontology. In: Proceedings of the Workshop on Semantic Search (SemSearch 2008), 5th European Semantic Web Conference (ESWC 2008), *CEUR Workshop Proceedings*, vol. 334, pp. 59–73 (2008)

85. O'Reilly, T.: What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software (2005). URL http://oreilly.com/web2/archive/what-is-web-20.html

86. Overell, S.: Geographic Information Retrieval: Classification, Disambiguation and Modelling. Ph.D. thesis, Imperial College London, Department of Computing (2009)

87. Overell, S., Rüger, S.: Using co-occurrence models for placename disambiguation. International Journal of Geographical Information Science **22**(3), 265–287 (2008)

88. Parundekar, R., Knoblock, C., Ambite, J.: Aligning Ontologies of Geospatial Linked Data. In: Workshop On Linked Spatiotemporal Data at the 6th International Conference on Geographic Information Science, GIScience 2010, *CEUR Workshop Proceedings*, vol. 691 (2010)

89. Polleres, A., Hogan, A., Harth, A., Decker, S.: Can we ever catch up with the Web? Semantic Web Journal **1**(1), 45–52 (2010)

90. Priedhorsky, R., Terveen, L.: The Computational Geowiki: What, Why, and How. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW 2008, pp. 267–276. ACM (2008)

91. Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man and Cybernetics **19**(1), 17–30 (1989)

92. Roget, P., Kirkpatrick, B.: Roget's Thesaurus. Penguin, London (1998)

93. Santos, D., Cardoso, N., Cabral, L.: How geographic was GikiCLEF?: a GIR-critical review. In: Proceedings of the 6th ACM Workshop on Geographic Information Retrieval, GIR'10, pp. 1–2. ACM (2010)

94. Santos, D., Chaves, M.: The place of place in geographical IR. In: 3rd Workshop on Geographic Information Retrieval, SIGIR, pp. 5–8. ACM (2006)

95. Schade, S., Maué, P.: Standardizing the Geospatial Semantic Web? In: Semantic Web meets Geopatial Applications Workshop, 11th International Conference on Geographic Information Science, AGILE 2008, pp. 1–5. Association Geographic Information Laboratories Europe (2008)

96. Schwering, A.: Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. Transactions in GIS **12**(1), 5–29 (2008)

97. Singh, P., Lin, T., Mueller, E., Lim, G., Perkins, T., Li Zhu, W.: Open Mind Common Sense: Knowledge acquisition from the general public. In: On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE, *LNCS*, vol. 2519, pp. 1223–1237. Springer (2002)

98. Smart, P., Jones, C., Twaroch, F.: Multi-source Toponym Data Integration and Mediation for a Meta-Gazetteer Service. In: Proceedings of the 6th International Conference on Geographic Information Science, GIScience 2010, *LNCS*, vol. 6292, pp. 234–248. Springer (2010)

99. Smith, B.: Ontology. In: L. Floridi (ed.) The Blackwell Guide to the Philosophy of Computing and Information, pp. 153–166. Blackwell, Oxford, UK (2003)

100. Staab, S., Gómez-Pérez, A., Daelemana, W., Reinberger, M., Noy, N.: Why evaluate ontology technologies? Because it works! Intelligent Systems, IEEE **19**(4), 74–81 (2004)

101. Strasunskas, D., Tomassen, S.: Empirical insights on a value of ontology quality in ontology-driven web search. In: On the Move to Meaningful Internet Systems: OTM 2008 Workshops, *LNCS*, vol. 5333, pp. 1319–1337. Springer (2008)
102. Strube, M., Ponzetto, S.: WikiRelate! Computing semantic relatedness using Wikipedia. In: Proceedings of the 21st National Conference on Artificial Intelligence, AAAI'06, vol. 2, pp. 1419–1424. AAAI (2006)
103. Suchanek, F., Kasneci, G., Weikum, G.: YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. In: Proceedings of the 16th International Conference on World Wide Web, pp. 697–706. ACM (2007)
104. Sui, D.: The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. Computers, Environment and Urban Systems 32(1), 1–5 (2008)
105. Tartir, S., Arpinar, I., Moore, M., Sheth, A., Aleman-Meza, B.: OntoQA: Metric-based ontology quality analysis. In: IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, at the 5th IEEE International Conference on Data Mining 2005, ICDM'05, pp. 1–9. IEEE (2005)
106. Turner, A.: Introduction to Neogeography. O'Reilly Media, Sebastopol, CA (2006)
107. Vaccari, L., Shvaiko, P., Marchese, M.: A geo-service semantic integration in spatial data infrastructures. International Journal of Spatial Data Infrastructures Research 4, 24–51 (2009)
108. Vaid, S., Jones, C., Joho, H., Sanderson, M.: Spatio-textual indexing for geographical search on the web. Advances in Spatial and Temporal Databases pp. 923–941 (2005)
109. Vander Wal, T.: Folksonomy (2007). URL http://vanderwal.net/folksonomy.html
110. Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., Studer, R.: Semantic Wikipedia. In: Proceedings of the 15th International Conference on World Wide Web, pp. 585–594. ACM (2006)
111. Weikum, G., Kasneci, G., Ramanath, M., Suchanek, F.: Database and information-retrieval methods for knowledge discovery. Communications of the ACM 52(4), 56–64 (2009)
112. Winter, S.: Ontology: buzzword or paradigm shift in GI science? International Journal of Geographical Information Science 15(7), 587–590 (2001)
113. Xiao, R., Liu, W., Du, Y., He, Y.: An intelligent bay geo-information retrieval approach based on Geo-ontology. In: Proceedings of the 17th International Conference on Geoinformatics 2009, pp. 1–6. IEEE (2009)