III. Communication by means of technology

16. Internet-Based Communication

Ulli Waltinger and Alexa Breuing

Today, internet-based communication encompasses a broad range of media services and technologies. While its architecture is still based on the traditional client-server model, its associated principles of *communication* and *information* are affected by the rising web-based communication technologies from the area of the *Web 2.0* and the *Semantic Web*. The current developments within the web have significantly stimulated many people to become involved in web-based communication. Accordingly, the web has changed from a medium for passive consumers to a network of active users who communicate via social networks, contribute to online communities, write weblogs, and publish podcasts. This survey chapter focuses on web-based technologies for human-human communication, with a special emphasis on rising technologies and applications. It reviews the principles of communication technologies within the *Web 2.0* and the *Semantic Web* and provides an overview of existing applications of recent web-based communication techniques.

1. Introduction

> *Computer Mediated Communication is a process of human communication via computers, involving people, situated in particular contexts, engaging in processes to shape media for a variety of purposes.* December (1996)

The internet has influenced and changed the way people communicate to an unprecedented degree. In modern times, online media technologies have become indispensable for the global communication process not only within the business area but also within our daily life. These developments in turn have significantly stimulated many users to become involved in web-based communication. In recent years, the World Wide Web has changed from a medium for passive consumers to a network of active users, who write weblogs (e.g. *Blogger*), participate in wikis (e.g. *Wikipedia*), publish podcasts (e.g. *iTunes*) or contribute and communicate within online communities such as using *Facebook* or *LinkedIn* (Stocker and Tochtermann, 2009). It has influenced the way we communicate (Mulvenna et al., 2000), interact (Kraut et al., 1998; Katz and Rice, 2002), shop, work, and inform ourselves (Jones, 1998). We might even argue that the internet helps to bring people closer together as it enables easy and cheap communication across vast distances (Gates, 2000). And although not everybody in the world possesses an internet access, the digital communication has become ea-

sier, faster, and, in principle, free. Altogether, the internet has obviously changed the way people collaborate and interact with each other, without regard to their geographic location.

The current web is characterized by its flexibility. Recent online publishing tools have simplified matters so that anyone can write, publish, and contribute data to the web. By now, internet *users* have become internet *editors* by taking a more active role in the collective development of a global knowledge (Surowiecki, 2004; Hammwöhner, 2007a). Thus, the internet has turned into a network of creative knowledge, where different cultures and disciplines have linked up to make something that is sometimes called the *collective intelligence* (O'Reilly, 2005). It establishes a mixture of very general and very specific interests and is thus a very precise, *self-limiting* and *self-organizing form* (e.g. in the domain of social software) (de Kerckhove et al., 2008). Hence, despite the development of web-based technologies, we have also witnessed a tremendous growth of data published on the web. However, as the available information increases, the inability of users to assimilate and profitably utilize such large amounts of information becomes increasingly apparent. For computer scientists, this body of information has opened up a new area of research, to develop and test new computational methods and to benefit from the collectively organized knowledge within the online medium (Mills, 2009).

> *Internet-based, computer-mediated communication involves information exchange that takes place on the global, cooperative collection of networks using the TCP/IP protocol suite and the client-server model for data communication. Messages may undergo a range of time and distribution manipulations and encode a variety of media types. The resulting information content exchanged can involve a wide range of symbols people use for communication.* December (1997)

People consider the WWW and the internet to be synonyms. From a technical standpoint, this is wrong. Initially, the internet refers to a system composed of computers connected by physical objects (e.g. cables). It is a physical network which facilitates the transfer and exchange of data. In this respect, the *ARPANET* (Advanced Research Projects Agency Network) was one of the first prototypes of such a system of communication (in 1969) (Abbate, 2000). It realized the first version of the *Transmission Control Protocol* (TCP) and the *Internet Protocol* (IP), used to exchange data and files and to connect to other computers (Caldarelli, 2007).

The WWW can be seen as a service layered over the internet. It refers to a hypertext system which is primarily based on three components: the *Hypertext Transfer Protocol* (HTTP) as the protocol to transfer data over a network; the *(eXtensible) HyperText Markup Language* (X)HTML allowing the definition of how the accessible information (web document) is structured, presented and connected to each other via hyperlinks; and finally, a *Uniform Resource Locator*

(URL) which serves as a unique identifier (web address) of a given resource (web document). Thereby, hyperlinks are not only used to interlink web documents but to interconnect other resources within the WWW as well. One of the first global hypertext projects was proposed by Tim Berners-Lee at the nuclear research center CERN in Geneva in 1989. Initially brought into being to share research findings and to improve collaboration with colleagues, in 1990 Berners-Lee presented the first web viewer that enabled users to view (browse) the first published texts. Subsequently, the web browser technology improved, starting from *Viola* (Pei Wei in 1992)[1], which added the ability to display graphics, to *Mosaic* (Marc Andreessen in 1993)[2] which turned into the later *Netscape Navigator* (1994)[3] and quickly became the leader in the rapid growth stage of the WWW (Abbate, 2000). In the 1990s, the WWW consisted of only a few static web pages which were manually listed in a first web directory/catalogue. But gradually, with the development of the first web search engines (e.g. *AltaVista*[4] and *Yahoo!*[5] in 1995, *Google*[6] in 1998) and decades after the first idea of a knowledge identification and retrieval system (*Memex*) (Bush, 1945), it soon led to the phenomenon that not only companies but also private persons tend to have individual *home pages* – accessible and findable on the web. Since that time, we have been able to observe a rising popularity in participation, contribution, and interaction with web-based technologies. This progress transformed the web into what nowadays is called *Web 2.0* (O'Reilly, 2005), *Semantic Web* (Berners-Lee et al., 2001), or *Social Semantic Web* (Blumauer and Pellegrini, 2009) – the web as we know it today.

2. Technologies of web-based communication

Technical developments in terms of an increase in broadband connections, a spread of internet access, and a decrease in prices of storage media (which are needed to archive such an enormous amount of digital data) was attended by significant developments regarding web-based communication technologies. While at the beginning of the WWW, say *Web 1.0*, it was all about getting people to publish and to find and connect information via hyperlinks, the WWW gradually transformed into a *web of social participation* (Mills, 2009). As a result, not only documents but also users are connected with each other. Since 2005, the term "Web 2.0" has become a popular buzzword for a number of interactive and collaborative elements within the WWW and we may argue that a so-called *Web 3.0* is currently emerging (Berners-Lee et al., 2001); a web where the focus is on the *automatic* identification and representation of the *semantic meaning* of information published on the web. More precisely, where the focus is on *connecting knowledge* rather than on documents and on putting them into a context.

Figure 1. Ideas that radiate out from the Web 2.0 core by O'Reilly (2005).

2.1. Communication within the Web 2.0

From a technical perspective, the phenomenon of Web 2.0 actually refers to techniques and services mostly developed in the late 90s. These include *Rich Site Summary* or *Really Simple Syndication* (RSS) (e.g. Netscape RSS in 1997) and web services such as *Application Programming Interfaces* (APIs), *Asynchronous JavaScript and XML* (Ajax in 1998), *Google* (1998), *Wikipedia* (2001) or *Flickr* (2002). Today, the term *Social Software* (referring to, for example, *wikis*, *weblogs*, and *microblogging services*) is mentioned as a special characteristic of the Web 2.0 (Blumauer and Pellegrini, 2009). So what is actually meant by "Web 2.0" and how does it facilitate web-based communication? In 2005, Tim O'Reilly (2005) presented the following initial sketch of the main principles of the Web 2.0 (see Figure 1):

1. *The Web as a platform:* In the beginning of the WWW, we could identify a clear separation between offline software applications and online web services. More precisely, in the past, software platforms and tools were merely installed locally on a single computer whereas web-based applications were mostly provided as *read-only* web services. Thus, even tools which were used to access and browse the WWW were based on offline software appli-

cations (e.g. Netscape Browser or AOL client). Thereby, all information and data gathered on the web had to be stored additionally on the user's hardware. Online bookmarks (in terms of favorite website hyperlinks), for example, could only be saved within the browser toolbar of the offline application. A true reciprocity or inter-connection between the local installed software tools and the different web resources was rather limited.

With the emergence of new online application services associated to the Web 2.0 such as the online photo-service *Flickr* or so-called social bookmarking tools such as *del.icio.us*, this paradigm changed. Users were now enabled to use online applications just like offline software. That is, they were enabled to contribute, store, access and supplementarily share their own created data and information within different online web services, always accessible from different interfaces (e.g. pc or cell phone) and independent from their current location. Accordingly, the users started to use the web as a new platform, just like their own computer. Meanwhile, it is common usage not to store individual documents, photos, or bookmarks on a single computer but additionally or even only on certain online service platforms and hence to share, categorize, and exchange this data with other internet users. A currently much discussed approach utilizing this trend is *cloud computing*. Cloud computing refers to both the applications delivered as web services and the hardware and system software that provide those services (Armbrust et al., 2009). By means of so-called *clouds*, people are able not only to provide data but also to access other computational resources such as software or calculating capacity by means of a computer network. Cloud computing especially has commercial relevance, as the flexible allocation and utilization of services as well as extensive capacities allow a great economy of scale (Baun et al., 2011). However, applications for private use receive substantial boost as well. Thus, for example, *Apple* already provides its customers with several cloud computing services (as *iCloud*). Altogether, we may argue that the vision of O'Reilly (2005), who stated in 2005 that the web is *ubiquitous and serves as a platform without hard boundaries*, is currently at the point of origin.

2. *Harnessing collective intelligence:* Any internet user may write and contribute content to the web and any other user is allowed to edit this content. Accordingly, using the web as a platform allows the harnessing of collective intelligence in terms of *Weblogs*, user-created structures such as *Folksonomies* or collaborative work by means of *Social Software* (see Section 3). This principle of harnessing *the wisdom of crowds* (Surowiecki, 2004) also forms the basic idea behind the collaborative approach for software development (see Principle 4).

   One aspect of collective intelligence becoming popular and possible by means of Web 2.0 technologies is *crowdsourcing*. It describes the act of out-

sourcing tasks through an open call to an undefined group of people and gathering those who are most fit to perform the announced task (Howe, 2008). Common examples are the online encyclopedia *Wikipedia* and *OpenStreetMap* where users can upload and edit vector data.

3. *Data as the next Intel Inside:* Database management is a core competence of most web companies as data present the main component of Web 2.0 applications. Every internet user accesses, utilizes, and often even extends the data on the web. Thus, the question of accessibility, usability, and ownership of this data needs to be clarified and put into the focus. Think of *Facebook*, where the data is made up of personal information provided by the numerous Facebook users. The more people participate and eventually assemble the data, the more is its increase in value. As a result of its currently more than 750 million active users[7], Facebook's value has, in the meantime, swelled to $50bn[8].

   In future, data suppliers and application vendors will battle for certain data classes that will become highly important for Web 2.0 applications. Companies already begin to realize that control over data may be their chief source of competitive advantage.
4. *End of the software release cycle:* Software is no longer distributed as a product/software release but as a service within the web. The implementation and testing of software become a public business turning internet users into co-developers. Thus, software development turns into a practice extending the open source philosophy *"Release early. Release often."* (Raymond, 2001) in that the product is additionally developed openly. Only by this, Web 2.0 applications are capable of meeting the challenges of daily operations and constantly changing user needs. Consequently, many open source initiatives (for example *Linux*) have obtained high quality results from collaborative work (Vivacqua and Borges, 2010). In addition, web developers have the opportunity to try out new features by providing them on their websites and monitoring the corresponding user behavior in real-time.
5. *Lightweight programming models:* Web services are more accepted and more prevalent the simpler they are. Thus, they have to be realized consistent with the principle *"innovation in assembly"* (O'Reilly, 2005). Therefore, lightweight programming models have to be supported in a way that allows an easy combining of content and features of different, existing web services to provide new services. The focus has to be on syndication rather than on coordination and on lowering the barrier to the public to (re-)use data and services. Furthermore, the service design needs to provide for "hackability" and "remixability" that enable and almost invite the harnessing of collective intelligence and thus to improve the web service collaboratively (O'Reilly, 2005).

Based on the principle of simplicity, many web services achieved great success. *RSS*, for example, has become perhaps the most widely deployed web service as it is simple, open, and easy to extend. Another example illustrating the great popularity of simple web services compared to complex ones is *Amazon* which provides its web services in two ways: On the one hand strictly adhering to the technical formalisms of *SOAP* (Simple Object Access Protocol) and, on the other hand, utilizing the simple distribution of *XML* data over HTTP based on a lightweight approach known as *REST* (Representational State Transfer). The majority of the Amazon usage (even 95 percent) is of the service building on REST (O'Reilly, 2005). This clear user preference is observable within other web services as well and it is not at all absurd to state that *Google* became the most successful search engine due to the simplicity of its application.

6. *Software above the level of single devices:* Web applications and services may no longer be limited to the PC platform but should also be suitable for mobile terminals and other devices. *iTunes* is the best example to date of the accomplishing of a seamless transition from several mobile devices to a massive web back-end from scratch. Other examples are *Skype* and *Facebook* which, amongst others, are available as single apps for smart phones. As a result, this Web 2.0 domain strongly drives the development of new, widely varying mobile devices such as iTV and traffic and activity monitoring platforms.
7. *Rich user experience:* The focus is on creating web-based software that works in a similar fashion and is perceived by the user also to be similar to a device-based software application. That is, on combining features of graphical user interfaces (e.g. drag-and-drop) and multimedia content on the web by means of client-side programmability and asynchronous data retrieval (e.g. *XMLHttpRequest* via Ajax (Garrett, 2005)). This enormous innovation in the field of user interfaces finally enables the construction of web applications which are as rich as local PC-based applications. Furthermore, we can expect to see many newly conceived applications as well as re-implementations of PC applications in the near future. Actually, *Gmail* has already provided innovations in e-mail which combine the strengths of the web with a browser-based user interface that is conform with independent e-mail software such as *Outlook Express* or *Thunderbird*.

Obviously, the majority of these principles can be positively attributed to current web interaction and communication techniques. The future address book will utilize web-based synchronization processes similar to Gmail that remember every mail, e-mail address, and phone number and which composes a separate social network allowing for a fast contacting and communication. Apple's *iCloud* already veers towards this evolution as synchronizing contacts,

calendars, and e-mails as well as music, apps, etc. between Macs, PCs, and iOS-devices via up- and downloads of data within a cloud. Furthermore, collaborative writing will reach a higher level as it will not be limited to editing only but will be supported by word processing applications similar to those for stand-alone documents (e.g. *Microsoft Word*).

The actual implementation of the Web 2.0 led to the development of so-called *personal publics* referring to online networks of communicative expressions regarding topics of mostly personal relevance (Schmidt, 2009). Such networks mainly include individual-related information such as one's date of birth, hobbies, or current situation and generally obtain rather small publics often consisting of already known people (see Section 3 for examples). Personal publics differ from other communication situations in terms of four attributes (cf. (Boyd, 2008)):

1. *Persistence:* Online expressions are recorded and available permanently. In contrast, information exchanged during everyday communication situations might be forgotten shortly after.
2. *Replicability:* As content is available in digital form, they can be copied without degradation and possibly even unnoticeably.
3. *Scalability:* Although personal publics might reach only a few audiences, their potential visibility is great. Many people, for instance, became prominent just by means of private videoclips provided online (e.g. on *Youtube*).
4. *Searchability:* Information can be accessed through search, uncoupled from their context.

In particular, Web 2.0 applications such as weblogs and social networks (see Section 3.4, 3.7) provide personal publics and thus generate another type of human communication based on technical web platforms.

2.2. Communication within the Semantic Web

The so-called "Semantic Web" can be seen as an extension of the current web (Web 2.0). It is sometimes referred to as a component of an upcoming Web 3.0 (Lassila and Hendler, 2007). In general, it is a vision that web resources should not only be readable and understandable for humans but also for machines, thus offering the crucial possibility of automation (Berners-Lee et al., 2001). The Semantic Web initiative has the objective to realize this vision by focusing on semantic annotations of web content and the usage of a shared terminology (see Figure 2). Thereby, the objective is not to establish a new web but rather to extend the existing WWW by a semantic layer.

> *The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.* Berners-Lee et al. (2001, pp. 2)

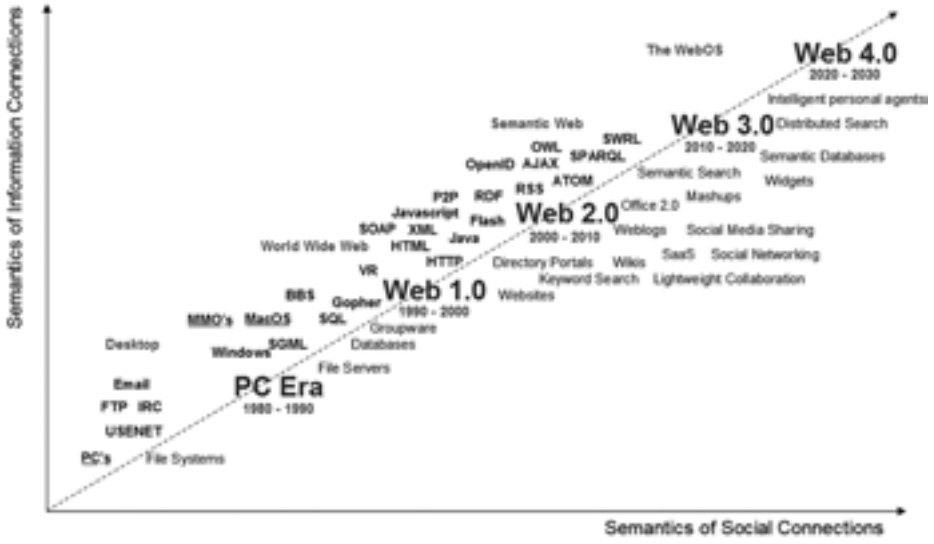

Figure 2. Ideas that radiate out from the Web 3.0 by Spivack (2007).

The semantic layer can be achieved by a more precise description of web documents by means of *metadata* (data about data) assigning semantics to the document content (Antoniou and van Harmelen, 2004). Hence, the Semantic Web vision follows the idea of connected intelligences in an ubiquitous web, where both humans and machines communicate with each other (*content intelligence*) (Mills, 2009). More precisely, in contrast to the Web 2.0, the Semantic Web is more about a web of *relations between resources* that denote real world objects such as people, countries, or events rather than a web of *accessible documents* (Guha et al., 2003).

> *The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. […] It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents.* Herman (2001, pp. 1)

Currently, the primary approach to achieve the annotation of documents is to define an ontology (e.g. *OpenCyc*[9], *GOLD*[10]) (see also Sharoff and Hartley (2012) in this volume). According to Gruber, an ontology is "*an explicit specification of a conceptualization*" (Gruber, 1993, pp. 199). The essential aspect of ontologies is a shared understanding of a specific domain among different web agents and applications and to subsequently use references to the concepts in the ontology as semantic annotations on top of web resources (top-down/bottom-up approach) (Berners-Lee et al., 2001).

> *Ontologies are the vocabulary and the formal specification of the vocabulary only, which can be used for expressing a knowledge base […] It should be stressed that one initial motivation for ontologies was achieving interoperability between multiple knowledge bases.* Hepp et al. (2008, pp. 6)

That is, users may create applications that collect web content from diverse sources, process, analyze, and annotate the information, and exchange their results with other applications by means of semantic markup specifications (Berners-Lee and Fischetti, 2000). Correspondingly, future software agents will carry out tasks for users rather than searching for information. Thus, completion of jobs such as finding a medical specialist nearby or setting up an appointment no longer consists of accomplishing numerous key-word-based search queries on the web and browsing the resulting text documents for important information manually. Instead, individual web agents communicate with other web agents, take into account one's circumstances such as the living place or time preferences, access one's personal calendar, and even reschedule appointments to present a solution. As much information exists only implicitly, the agents need a set of inference rules that enables them to conduct automatic reasoning (Berners-Lee et al., 2001). Hence, for Berners-Lee and colleagues, the challenge of the Semantic Web is to provide a representation formalism (ontology) that allows for reasoning about data to infer information which is not available explicitly (e.g. the distance from one place to another or connections between two persons). Currently, the most widely accepted languages for the description of ontologies are based on the *Resource Description Framework* (RDF) (Manola and Miller, 2004; Hayes, 2004; Klyne and Carroll, 2004). RDF is a markup language recommendation proposed by the *World Wide Web Consortium* (W3C) for the description of resources and their properties in a machine-processable manner and it provides the basis for most ontology languages (May, 2006). The identification of resources is ensured based on so-called *Uniform Resource Identifiers* (URIs) assigning a clear description to any resource on the web. The most familiar type of an URI is the *Uniform Resource Locator* (URL) which specifies where a specific resource is available and how to retrieve it. In contrast, URIs are more general and can refer to everything, thus additionally to objects of the real world such as persons, institutions, or words (Walton, 2006). RDF possesses a vocabulary containing a particular set of URIs with predefined meanings (Hayes, 2004; Hitzler et al., 2009). Additional information on URIs is available within the corresponding RFC (Berners-Lee, 1994).

RDF enables both the allocation of information in a human and machine-readable format and the realization of efficient data exchange. Thereby the resource description may be available as a data model in terms of a graph or serialized in a particular syntax. The former allows the relation of resources to other resources by means of triple-based *statements* each consisting of a subject (node), a predicate (link), and an object (node) (see Figure 3). As the object may

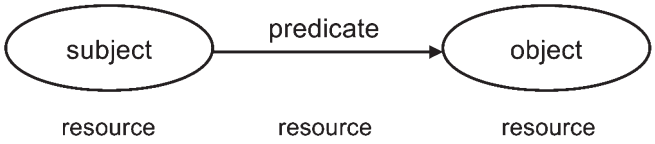

Figure 3. Components of a RDF statement.

be the subject of further statements, several statements can be concatenated and thus build up a directed graph (Walton, 2006). The serialized representation of RDF is based on *RDF/XML* which benefits from the advantages provided by *XML* (simplicity, generality and usability) but additionally enables the description of machine-understandable semantics (Beckett, 2004) (see also Stuhrenberg (2012) in this volume). The data is often stored in a relational database or in a database optimized for the storage of statements called triplestore. Altogether, RDF is a general-purpose language for the representation of semantic information on the web (Brickley and Guha, 2004). However, to build up ontologies, resources need to be arranged in categories enabling the definition of statements about classes of resources (Birkenbihl, 2006).

RDF Schema (RDFS) (Brickley and Guha, 2004) is an extension of RDF in that it provides basic elements for the definition of ontologies. In RDFS, resources with the same properties are considered to be instances within one class. Classes in turn are resources which can be referenced via URIs and which may have own properties. RDFS supports both the definition of sub-class relations and the definition of completely new relations (Aitken et al., 2004). The latter can be limited by domain and range specifications. Thus, RDFS constitutes an ontology language that allows for the definition of vocabularies, the specification of properties and their values in connection with different classes, and the definition of relations between resources (Stuckenschmidt and van Harmelen, 2005). However, the inability to define equalities between resources and additional limitations regarding properties present a substantial deficit of RDFS. Then again, exactly this was precisely a consciously made design decision to possess a simple and plain language for ontologies (Walton, 2006).

The mentioned inadequateness of RDFS is remedied by the allocation of a more expressive ontology language. The *Web Ontology Language* (OWL) (McGuinness and van Harmelen, 2004; Patel-Schneider et al., 2001, 2004) (see also Stührenberg and Sasaki and Lommel both in this volume) can be used to represent ontologies just like RDFS, but additionally provides more possibilities to express content and meanings. Thus, OWL extend RDFS by classes and properties allowing the definition of new classes based on enumerations of instances, calculations of intersections and unifications as well as specifications of cardinalities and complementary operations (Birkenbihl, 2006). However, a higher expressiveness always goes hand in hand with a reduced support for ef-

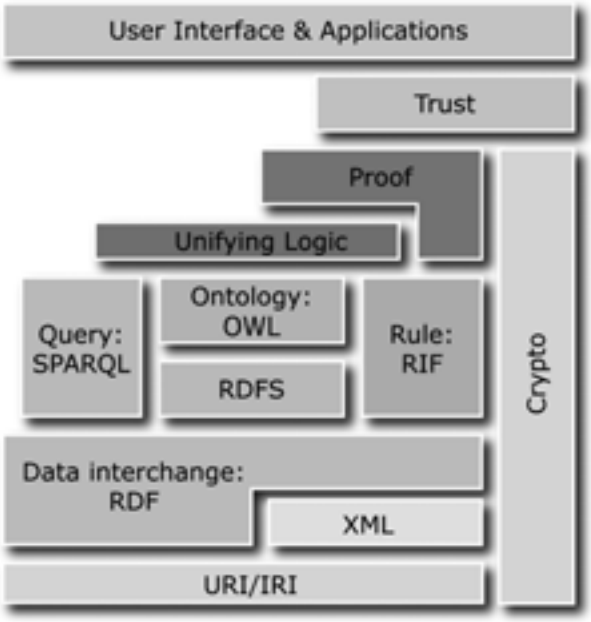

Figure 4. The Semantic Web Layer Cake proposed by (W3C, 2009).

ficient deduction. For this reason, three different levels of languages were defined: *OWL Lite*, *OWL DL*, and *OWL Full*. The expressiveness increases with each level, so that OWL Full is the most expressive one (Walton, 2006). Stuckenschmidt and van Harmelen (2005) define the three levels as follows:

1. *OWL Lite:* supports classification hierarchies and allows the formulation of equalities, inequalities, and simple constraints (e.g. cardinality values of 0 or 1).
2. *OWL DL:* allows both maximum expressiveness and computational completeness as well as decidability at the same time. "DL" refers to the correspondence with description logic, a knowledge representation language used for formal reasoning on terminological knowledge.
3. *OWL Full:* allows maximum expressiveness and syntactical liberty but is not determinable.

The manual creation of semantic annotation is both an expensive and time-consuming task that is extensively studied in the area of learning and managing ontologies (Staab and Studer, 2004). In this context, a lot of work has been done on developing tools for ontology engineering in order to support users in the process of creating ontologies (see, e.g., the *Protégé Ontology Editor and Knowledge Acquisition System* (Knublauch et al., 2004) or the *SWOOP* project

(Kalyanpur et al., 2004)). Miller (2001) from the W3C (W3C, 2009) describes the guiding architectural principles (see Figure 4) of the Semantic Web as follows:

1. *Everything Identifiable is on the Semantic Web:* Referring to the Semantic Web, all resources on the web need to have an identity. That is, people, countries, and other data objects referred to need to have an online representation in the form of an RDF-URI. URIs should support the effective integration, participation, and contextualization of data within the Semantic Web (Manola and Miller, 2004).
2. *Partial Information:* In accordance with the idea of the *web as a platform* (see Section 2.1.1), the web is ubiquitous in that it lacks hard boundaries. Semantic Web technologies need to focus on the aspect to *enable web access for anyone, anywhere, anytime, using any device* (Hoschka, 2011). In addition, the Semantic Web should tolerate partial information, such as missing links (e.g. non-existing targets) and/or incomplete or inconsistent information. Every user may contribute data and information to the web by creating different types of links to (web) resources (e.g., RDF-URI), which can be discovered and/or differently typed by others.
3. *Web of Trust:* The web is a collaborative environment (not read-only) that should offer confidentiality and confidence but simultaneously allows users to take responsibility for what they publish on the web (e.g. via signatures, group authoring, or versioning). In addition, obviously, all data, information, and statements accessible on the web occur in a specific context. It is the job of the respective applications to check the reliability and to evaluate the trustworthiness of the information found on the web.
4. *Evolution:* The Semantic Web should support the independence of different communities. According to their own specifications, these communities should be able to add and distribute information independently of other standards. Thereby it should be based on some *descriptive conventions* which can also be understood by the users.
5. *Minimalist Design:* In the context of the Semantic Web the *principle of minimalist design* refers to the aspect that every formulation of a *semantic assertion* should be based on a common and universally valid model (Berners-Lee, 1998). The idea behind is that it is the general validity and simplicity of a model which enables other applications to use, link or map to the formulated assertions.
6. *Building on Proven Ideas:* The Semantic Web community provides already existing technologies, formats, and frameworks to annotate and interlink data objects (e.g. RDF, OWL). Emergent applications and technologies should agree to the general principles of web automation, data aggregation, and support the establishment of a common syntax and vocabulary (Stumme et al.,

2011). For instance, to provide simple applications that do not ignore the possibility of additional complexity in the future (e.g. *Dublin Core* (Dublin Core Metadata Initiative, 2010) – see also Trippel (2012) in this volume).

The goal of the Semantic Web is to convert the web into a distributed, machine-understandable data resource by means of semantic markup specifications and metadata. This metadata should be defined, formulated, and structured in a way that, at best, enables machines and applications not only to process but also to understand it correctly (in terms of reasoning). Furthermore, it should allow for drawing conclusions based on existing data as well as the generation of new information in this way. Data objects within the Semantic Web do not target its direct usage by people but by machines. However, this approach implies several challenges for the computer science domain: Due to the enormous amount of data and information available on the Web, *it is impossible to manually enrich all of these resources* (Stumme et al., 2011). Furthermore, one faces questions of the following sort: How to establish such a semantic backbone for such a large number of distributed web resources? How to query, process, and analyze such an amount of ontology and *RDF* data efficiently on a decentralized network (see approaches like *SPARQL* (Prud'hommeaux and Seaborne, 2008))? How to ensure that two different web agents share the same semantics for a given annotated concept (in terms of standardizing the information distribution across the network)? In essence, there is the need to automatically learn, interpret, organize, and retrieve relevant information and resources so as to meet the heterogeneous requirements of users and agents (Chishti et al., 2009; Stumme et al., 2011).

3. Applications of web-based communication

A variety of web tools, applications, and (social) resource-sharing systems and networks, such as *wikis*[11], *weblogs*[12], *Twitter*[13], *Facebook*[14], *del.icio.us*[15] or *Flickr*[16]. are associated with the current web (Cattuto et al., 2007). Obviously, these mainly user-centric publishing and knowledge-management platforms have acquired a large number of users in recent years. One of the main reasons for the success of these platforms is that there is no specific competence needed in order to participate. Built-in tools enable even non-technicians to contribute to the web and to engage in web-based social communities.

At first, *e-mail* and *newsgroup communication* had taken a major role in internet-based communication. Subsequently, the first online-based *bulletin boards* and *guest books* provided an additional platform where users could communicate via comment functions and forum posts. However, until then, the actual content production and publishing process was left to commercial internet portals or online newspaper publishers. With the emergence of *weblogs*, the

first independent content publishers became present on the web. At that time, the term *user-generated content* was coined. Nowadays, telephoning, video messaging, and other services are also attributed to a *web-based communication* which, as we may argue, formed a second functional layer of interpersonal communication. Finally, with the rising popularity of social networks such as *Facebook* and *Twitter*, users not only became able to publish content on the web but also to aggregate and conjoin different web-based communication techniques with each other. More precisely, we can identify an amalgamation of different communication channels and online services now. Text messages sent by cell phones via the online service *Twitter* are indexed and searchable by *Google*; videos published by *YouTube* are interlinked to *weblogs*, rated and commented on *Facebook* and shown on *television* at home. Newly emerging web services within the Semantic Web support the connectivity of data by enabling the exchange of addresses, events, or social relationships (Brickley and Miller, 2010). That is, information (e.g. comments, images, or videos) which is published on one web site by a single user which may uses different technical devices (e.g. telephone, TV, or computer) may promptly appear in many other places (and hence in databases) on the web as well. This phenomenon relates to some crucial aspects of the Semantic Web (e.g. partial information – the web of data) but also raises privacy issues and the role of remembering in a *web that never forgets* (Mayer-Schonberger, 2009).

3.1. E-mail

By far the most common and oldest form of internet-based communication is *electronic mail* (e-mail).[17] In general, e-mail refers to the transmission of messages over communication networks. Messages and attachments (e.g. files and documents) can be sent using an offline e-mail client (e.g. *Thunderbird*) or an online web service (e.g. *Gmail*) through an internet gateway as provided by an *Internet Service Provider*. Today, sending e-mails is still a fast and effective way to synchronize real-time communications (Schuff et al., 2007). However, with the increase of e-mail users (more users sending more e-mails), information overload has become a significant problem in recent years. For example, it led to an increase of so-called spam e-mails (Youn and McLeod, 2007). Spam refers to an electronic advertising mail (junk) which is sent (automatically) to thousands of people. Researchers interested in e-mail have addressed these concerns with classification systems focusing on spam protection (Hayati and Potdar, 2008), relevance rating, and summarization (Aery and Chakravarthy, 2005; Xiang et al., 2007) as well as information extraction and disambiguation (Minkov et al., 2006) (see also Paaß (2012) in this volume). Finally, these systems also address e-mail-act categorization (e.g. request, commitment) (Carvalho and Cohen, 2005). Most recently, the search engine company *Google* presented

a web service called *Google Wave*[18] which aims to amalgamate the traditional (asynchronous) e-mail communication with collaborative elements of the Web 2.0 in real-time (e.g. instant messaging, wikis, and weblogs) within one enhanced e-mail application. This communication and collaboration tool enables users to exchange messages (so-called *waves*) as they would do it in e-mail or instant messaging conversations. The *wave* can contain text, comments, pictures, or embedded videos (e.g. *YouTube*, *Google Maps*). All *wave* collaborators work from a single real time copy. All replies and edits are visible in real-time for the collaborators (e.g. *letter-by-letter*). The *wave* represents in this context both a conversation and a document, it is equitable treated. This service was one of the first attempts that combined e-mail with instant messaging, documents with conversations, and allowed the web-based collaboration and communication within one *wave* object.

3.2. Instant messaging and video conferencing

Instant Messaging (IM) services are still one of the most common and fastest-growing forms of web-based communication. IM enables users to communicate (chat) with other users (chat groups) via text messages in real-time. Messages are sent via the network using a (standardized) messaging protocol such as *XMPP* (Extensible Messaging and Presence Protocol) or *OSCAR* (Open System for Communication in Real-time). In general, users need to have a chat client installed on a device (e.g. computer or cell phone) which is connected to a server via a network. Most of the IM-clients offer a function which alerts the user whenever a recognized contact within the user's buddy list (see Section 3.3.7) is online. One distinguishes between three main forms of instant/chat messaging forms (Storrer, 2001):

1. Traditional IM services are usually not listed in a public domain (e.g. chat rooms) and use the client-server or a peer-to-peer model for the communication (in terms of data transmission). That is, participants need to have the client software installed on their computer.
2. *Internet Relay Chat* (IRC) also requires its own chat servers, which are usually connected through a network. For the communication within the IRC client specific IRC commands are used.
3. Recently, web-based chat services have gained popularity, which enable communication/chat directly using the web browser as client. In contrast to IM or IRC, these services do not require any additional software but are mostly limited to a particular website (e.g. *GoogleTalk*, *Facebook*).

In this context, IM and chat services effectively combine features of the telephone (one-to-one near-synchronous interaction), e-mail (text-based communication), and chat-rooms (one-to-many communication). Altogether, these fea-

tures contribute in making IM inexpensive compared to other forms of media such as the telephone. Due to its convenience and usefulness and as a result to an increase in internet connection time, IM has also become popular in the business world (Glass and Li, 2010). Popular IM clients and services include *QQ*, *ICQ*, *AOL Instant Messenger*, *Microsoft MSN Messenger*, and IBM's *Lotus Instant Messaging*.

While IM enables users to communicate by means of text-messages, *Video Conferencing* (VC) enables two or more participants to communicate via audio and video data at the same time. That is, VC provides a visual interaction that cannot be achieved by the traditional IM or e-mail communication. In addition, most of the current VC clients integrate services for instant messaging (such as, for example, *Windows Live Messenger*). Research on IM and VC focuses on usability issues by means of message composition (by analyzing, for example, keystroke data – see Campbell 2004), the usage of IM for personal and business purposes with respect to conversational characteristics and functions of IM (Isaacs et al., 2002), on VC-based traffic classification (Angevine and Zincir-Heywood, 2008) and on linguistic characteristics of chat communication (Storrer, 2001; Beißwenger, 2007).

3.3. Internet telephony and VoIP

Another internet-based communication technique that has gained popularity recently is internet telephony, also known as *Voice-over-IP* (VoIP). VoIP refers to a method of transmitting voice data in packets over the internet (IP network) rather than by traditional circuit-switched telephone networks (Public Switched Telephone Network). Thereby, the voice data is converted into data packets, sent through the network, and converted back into voice data at the corresponding destination. If a telephone is used as a communication device, it needs to be connected to a VoIP phone adapter that in turn is connected to the network. Similar to an internet service provider, this technique is enabled by a VoIP provider such as, for instance, *Skype* or *Sipgate*[19]. In general, VoIP is less expensive than traditional voice communication. Hence, gratis VoIP services are already available in terms of device-independent lightweight applications (so-called *apps*, (see Section 2.1)). According to this and by means of an appropriate device, it is possible to chat with a relative living 1,000 miles away at no cost while being on the playground with one's children. In connection with VC, one can simultaneously present some livestreams from the playground to the relative to show how the children have grown up. As a result of the development of latest Web 2.0 techniques, such scenarios are commonplace today. However, the quality of the conversation via VoIP still depends on the available internet bandwidth (Bhanu et al., 2010). Compared to the conventional telephony, VoIP shows several advantages for both providers and users. In terms of the pro-

vider's part, internet telephony offers more developmental freedom as it is provided as a service, independent from old protocols or hardware. Thus, VoIP providers have more opportunities to exploit the new technology to fulfill the users' needs (Tapio, 2005). In terms of the users, the search for an adequate provider is not longer limited to those which provide their services locally and to those having roaming agreements with local operators in other areas. Another advantage for the users of quite a different nature is the increased control regarding their availability for conversations. In *Skype*, for example, users are able to specify their availability by means of icons appearing in the friend's buddy list (see Section 3.3.7). These icons symbolize if the corresponding contact is working, absent, etc., or if he/she is available. Despite of this information, many users additionally check for the availability of their contacts by sending them a message via IM as most VoIP applications involve this function as well. As a consequence, unannounced calls are increasingly regarded as annoyances (Joisten, 2007). Hence, interaction rules of IM and telephony have converged and generated a new form of multimedia-based communication application.

3.4. Weblogs

In recent years, weblogs have become rapidly popular as a new and easily accessible tool for *Personal Publishing* that is suitable even for non-technical users. In general, weblogs or *blogs* refer to (personal) websites where users publish (short) articles (sometimes referred to as *microcontent*) about their lives, opinions, or about technical issues (Blood, 2002, 6). It can be seen as a form of a web-based diary that is publicly available. The authors of weblogs (so-called *bloggers*) primarily use weblog-publishing-systems (e.g. *WordPress*[20]) or *Content Management Systems* (CMS)[21] to publish their content (*user-generated content*) (Vakali and Pallis, 2003; Stocker and Tochtermann, 2009). These systems provide the technical basis for the publishing process. At the beginning of the Web 2.0, weblogs gained much popularity by the online community, since these systems were free of charge, interactive, and easy-to-use. That is, there is no technical or web-design knowledge needed to setup a weblog system. Most weblog systems allow other users to post comments to the respective articles. Related weblogs, online friends, or bookmarks are linked together through the so-called *blogroll* section within the website. From a technical point of view, weblogs introduced several interactive elements which quickly helped them to become popular:

1. *Permalinks* refer to a permanent accessible URL linked to an item of microcontent (e.g. a single article or video) within the weblog. That is, single contributions are directly and permanently accessible regardless of where the content is located within the weblog.

2. *Trackbacks* or *Pingbacks* (SixApart, 2002) enable bloggers to determine whether another weblog refers to one of their own weblog entries. This feature sends (mostly automatically) a network signal (a so-called *ping*) to a specific URL of the target weblog, which in turn creates a backlink to the initial website. In this way, the weblog software is notified of new referenced weblog entries.
3. *RSS-Feeds* provide (possibly summarized) text including metadata such as publishing date, authorship, and backlink of the most recent entries of a weblog in a machine-readable XML format (e.g. based on *Really Simple Syndication 2.0* or *Atom Syndication Format*). The majority of modern weblog software systems automatically provide RSS for content syndication.

In the past, all these characteristics contributed to the creation of a virtual and interactive *weblog network*, which is sometimes called the *blogosphere* (Stocker and Tochtermann, 2009). In this context, weblogs have influenced the nature of web-based communication. They can be regarded as less objective due to their emotional communication style (Back et al., 2008, 18), have a greater diversity of topics covered, and enable users to engage in conversations on most recent issues almost in real-time. Consequently, the emergence of such user-generated content offers a wide range of applications not only to computer science (e.g. sentiment analysis (Melville et al., 2009), web genre classification (Santini et al., 2010) and topic detection (Sun et al., 2007)) but also for political science (Drezner and Farrell, 2008; Farrell and Drezner, 2008) and social psychology (Herring et al., 2004; Miura and Yamashita, 2007). It allows researchers to analyze what people think of and why (Kolari et al., 2006).

3.5. Folksonomy

Composed of the terms "folks" and "taxonomy", *folksonomies* (Wal, 2005) provide systems for the collaborative categorization of web content by allowing anyone, especially consumers, to annotate this content with arbitrary tags (Alby, 2008). They are used for tagging web pages (e.g. *del.icio.us*), annotating pictures (e.g. *Flickr*) or classifying scholarly publications (e.g. *bibsonomy*[22]) (Damme et al., 2007). Thereby, different kinds of tags can be specified, for instance tags that identify what (or who) the content is about, what is presented or who owns the content (Golder and Huberman, 2006). Thus, tagging features reflect subjective assignments between words and categories of objects, intersubjective patterns in these associations and implicit information on social networks (Damme et al., 2007). The advantage of folksonomies, compared with traditional taxonomies, is an intuitive, user-based annotation of data without predetermined concepts which allows a faster and simpler usage (provided by, e.g. *del.icio.us*). Typically, these sites enable users to publicly tag and share

content, that is, they do not only categorize information for themselves and they can also browse the content categorized by other users (Golder and Huberman, 2006).

Besides web sites that pursue this 'social' aspect of collaborative tagging, approaches aiming for commercial purposes only exist. *Google*, for instance, published a tagging game (*Google Image Labeler*[23]) where two people see the same image and label this image with as many tags as possible. To receive points (the goal of the game) both participants need to enter the same label. The more specific this label is, the more points can be awarded. By this, the quality of Google's image search results is improved by means of collective intelligence (see Section 2.1.2).

3.6. WikiWikiWeb

The *WikiWikiWeb* or *wikis* refer to the domain of *open collaboration hypertext systems*. Even though the first version of a wiki was published in 1995 (Leuf and Cunningham, 2001), wikis became first popular as a part of the Web 2.0 in 2001, namely by means of the rise of the online encyclopedia *Wikipedia*. The basic concept of a wiki is that users are allowed to not only read, write, publish, and contribute articles, but also to edit and (re-)structure existing documents of other contributors within the same wiki. That is, the content of a wiki is created and revised by a multitude of users in a collaborative manner. This principle follows also the notion of what Berners-Lee described as the *read-and-write web* (Berners-Lee and Fischetti, 2000). A well-known and widely used open-source wiki software is *MediaWiki* (e.g. used by the *Wikipedia* project[24]). Like traditional content management systems, the *MediaWiki* software allows even technically non-experienced users to publish content (e.g. texts and pictures) and to contribute to the corresponding community of *wikilocutors* in this way. That is, users do not need to have a sophisticated knowledge about markup languages (e.g. HTML) but only about the simple syntax of the wiki in order to publish data. Every article within a wiki corresponds to a separate page (assigned by an URL). Furthermore, users can attach cross-references to other pages via hyperlinks to express some kind of relationship (e.g. specification or definition). In this context, wikis establish a mix of very general and very specific interests and hence integrate a highly collaborative but also self-organizational property within the community (de Kerckhove et al., 2008). The enormous amount of documents available of the *Wikipedia* project has also drawn much attention within the scientific research community in the last years. As for instance, with regards to the quality of Wikipedia articles (Hammwöhner, 2007b), the relationship of Wikipedia categories (Chernov et al., 2006; Gleim et al., 2007; Waltinger and Mehler, 2009), its graph and hyperlink topology (Voss, 2005; Krötzsch et al., 2005; Nakayama et al., 2008; Mehler, 2006; Milne and Witten,

2008; Mehler, 2011), in the context of domain specific thesauri (Milne et al., 2006; Zesch et al., 2008) or semantic relatedness (Strube and Paolo-Ponzetto, 2006; Gabrilovich and Markovitch, 2007; Waltinger et al., 2009) In recent years, *Semantic Wikis* (Schaffert et al., 2008) – as an extension of the wiki idea – have drawn some attention in the research community. These software tools aim to combine the strengths of the Semantic Web (data integration and the realization of machine processing and complex search queries) and the *MediaWiki* technologies (easy-to-use, interactive, and emphasizing collaboration). That is, users can not only contribute web content but are also able to define and annotate entities and concepts within contributed articles. These annotations (possibly based on RDF – see Section 2.2) may refer to predefined or newly created types of relations in ontologies. For instance, the relation "has birthplace" may be used to connect a proper name with an article about a geographic location. The collaboratively organized annotation of metadata facilitates the automatic content analysis by enabling, for example, complex search queries using the *SPARQL* query language for RDF (as done, for example, in the *DBpedia* project – see Auer et al. 2007). In this context, *Semantic Wikis* can not only be used to improve the browsing and search experience (e.g. faceted browsing), but may also enable non-technical users to contribute structured information to the Semantic Web.

3.7. Social Networks

Today, the web is characterized by a variety of social networks such as *Facebook*[25], *Twitter*[26], *MySpace*[27], or *LinkedIn*[28] (see also Diewald (2012) in this volume). Unlike wikis or weblogs where the main objectives are the publishing process and the linkage of documents, social networks primarily focus on the self-portrayal of users and the establishment of social relations among them. That is, social networks allow people to not only present themselves on the web via user profiles, but also to explore, interact, communicate, and connect to other profiles (people) within the shared community. The members of a social network even interact and collaborate with people they did not know before, just by following acquaintance links starting from the profile of a neighboring acquaintance. However, most friends and communication partners are already known persons. Hence, social networks support the practice of *friendship-driven participation* (Ito et al., 2009). The basic principles within social networks are the following:

1. *User profiles* represent the nodes of a social network. Most network sites enable users to represent themselves by means of an online profile or portfolio that allows them to add and upload textual information (e.g. profession, gender or avocation), images, audio, and video files or documents.

2. *Buddy lists* represent the edges between the user profiles. That is, approved friends are added to a list of contacts or to a so-called buddy list. People finding themselves in these lists are allowed to post comments, view the user profile, and to forward the status of a friend to other users.
3. *Private messaging* (in terms of web-based e-mail messaging) is mostly enabled between approved online relationships only. Sophisticated online networks also provide communication (restricted to friends only) via instant messaging (chat) services.
4. *Bulletin boards* allow friends as well as the profile owner herself to leave marks in terms of textual messages, images, or videos that are readable and accessible for any other friend.
5. *Notifications* indicate whether a user has changed or added profile-related information (e.g. text messages, images or comments). Friends are often encouraged to post comments on the latest status update of an user. These status update features are sometimes referred to as *microblogging* features.
6. *Virtual groups* can be built by any member of the social network. That is, any user may create a group around a (common or shared) topic of interest and invite other members of the community to participate.

Communicating via bulletin boards has become very popular as it supports easy and fast spread of news, notifications, expressions of opinions, and public discussions. As a result, these boards obtain communicative power and social networks such as *Facebook* more and more become a counterpublics instrument (see the "facebook revolution" in Tunisia and Egypt (Hauslohner, 2011)). How easy and especially rapid publications are spread over social networks can be exemplified by the news of an unintentional invitation of 1,600 people to a party published worldwide: A student from Hamburg posted an invitation to her birthday party, which included her address, on *Facebook*, unfortunately without marking the event as "private". As each user who responds to a public event automatically invites his friends as well, thousands of Facebook users felt invited. In the end, 1,600 of them crashed the girl's birthday party (Bird, 2011). Most recently, a new form of social network has gained popularity within the online community. *Microblogging platforms* such as *Twitter* or *Tumblr*[29] differ from traditional weblogs and social networks in that they primarily focus on *status update features*. Here, published content (e.g. posts or tweets) is typically smaller than weblog articles as the published information only consists of a short sentence, comment, quotation, image, or video (e.g. the number of characters allowed at *Twitter* is 140). Addressed topics may range from small comments about recent issues (e.g. political statements, reviews of movies, or forwarded tweet messages, so called *retweets*) to what the user is currently doing or where he/she is now (e.g. geographical information) (Lenhart and Fox, 2009). However, studies of communication behaviors on *Twitter* show that the

exchange of reports, episodes, and anecdotes about the everyday life, that is, the exchange of personal information, predominate (Java et al., 2007b). These platforms often serve as modern news media spreading information rapidly, easily, and globally. Thus, the news about the first nuclear meltdown in Fukushima and subsequent news about evacuations, explosions, etc. were notified promptly on *Twitter* and other microblogging platforms, while this information was published in the daily newspaper two days later. The data provided by social networks has opened up new areas for research including political science (Mustafaraj and Metaxas, 2010), usage and intention (Bohringer and Barnes, 2009), user influence (Cha et al., 2010), microblogging spam (Lee et al., 2010), and opinion mining (O'Connor et al., 2010) (for text mining see Paaß (2012) in this volume). This also includes research on topological and geographical properties of social networks (Java et al., 2007a; Backstrom et al., 2010).

4. Challenges of web-based communication

With the sustained technical advances of web-based communication technologies and the rising number of online communities and participants, the WWW is gaining a growing importance for studies on computer-mediated communication (e.g. user-to-user) and human-computer interaction (e.g. computer (agents)-to-user) (Thurlow et al., 2004). Due to the convergence of emergent communication and interaction technologies (in terms of web-based communication via conventional technologies), a wide range of (linguistically motivated) heterogeneous data is associated with the web (Hancock et al., 2005). Modern communication technologies have enhanced the quantity of written or spoken data that manifests rich sources of collaboratively organized knowledge (of a widespread thematic diversity regarding subject-specific as well as common-sense knowledge). This data can be explored to improve information retrieval systems (Waltinger and Mehler, 2009; Mehler and Waltinger, 2009), but also to speed up information processing among users (Wilson et al., 2009; Canali et al., 2011). That is, newly emerging issues and events are published, spread, communicated, and also critically rated almost in real-time (e.g. think of presidential elections or news of natural disasters communicated via *Twitter*).

Today, the web can not only be seen as a mirror of language use within different linguistic and social communities, but also as a communication medium in which language use is shaped (Erbach, 2004). It enables researchers to explore the social cues on how people use and interact with their computers, with each other (in terms of network relations which are characterized by content and direction), or within groups (in terms of network ties between networked users) (Garton et al., 1997). Within this *Web of Data* (Berners-Lee et al., 2001), hyperlinks are more reflective of communication pathways between communities

than of traditional communication channels. In this sense, a hyperlink serves as an *indicator* (Park and Thelwall, 2003) of a *communication relationship* (Erbach, 2004). In addition, the web is currently changing from a web of *hyperlinked documents* to a web of *typed hyperlinked data* (Ding et al., 2008). Web-based communication as well as interaction services and applications such as *semantic wikis* (Schaffert et al., 2008), *social bookmarking* (Hotho et al., 2006) (in terms of *social tagging* (Mika, 2004; Staab et al., 2005)), and traditional wikis (in terms of *social ontologies* (Cattuto et al., 2007; Mehler, 2008, 2011)) contributed significantly to this phenomenon. However, due to the convergence of different computer-mediated techniques, the domain of web-based communication now also incorporates a broad spectrum of heterogeneous sources and data formats, which pose new challenges for the research community. That is, tools and methods developed for the automatic annotation of linguistic data can not be directly applied to computer-mediated data (e.g. chat annotations comprise non-standard spellings, conceptually oral forms, or written dialects) without being adapted (Beißwenger and Storrer, 2010). Therefore, the linguistic analysis of web-based communication must consider that language use within the web (say *netspeak*) differs from traditional linguistic data (e.g. *TIGER Treebank* (Brants et al., 2002), *Reuters Corpus* (Rose et al., 2002)) so that the tools and linguistic services must be extended to new tasks (Beißwenger and Storrer, 2010).

Another challenge is the control of privacy within a social web that makes topics and information of personal relevance publicly available (Schmidt, 2009). For instance, the access to information about persons situated in different social contexts causes role conflicts as several self-portrayals arise from a role-specific context but are accessible for persons possibly belonging to other reference groups (*unintended public* (Schmidt, 2009)). According to this, many employees do not want to be associated to their private self-portrayals within their job-related context. In reality, many recruiter search social networks for profiles of job candidates as expecting to figure out their true personality. Furthermore, internet users have to face the problem that other persons provide photos or information on the web regardless of whether they agree to this or not. Social psychological studies have shown inconsistencies regarding the relation between attitudes towards privacy and the behavior within the web. Thus, people having little interest in disclosing information do often not attach great importance to the privacy protection of others – also known as *voyeurism* (Schmidt, 2009). One way to retain the control of privacy is to present oneself variously for different publics (e.g. by creating self-portrayals on occupational networks like *XING*[30] and on a platform designed for personal, recreational interactions). Additionally, many platforms offer further mechanisms for privacy control by means of password protection or restricted access. However, especially due to the mentioned attributes of personal publics (persistence, replicability, scalabil-

ity, searchability; see Section 2.1) versant borders between the public and privacy perish and need to be renegotiated in terms of the current web.

5. Conclusion

Today, the area of internet-based communication encompasses a broad range of media services and technologies. Unlike previously, we can now identify an amalgamation of different communication systems and applications. Communication technologies have been transformed from being highly technical and specialized to being easy-to-use and personal – an integral part of our daily life. The developed technologies from the area of the Web 2.0 and the Semantic Web have made a significant contribution to these advances. With the tremendous growth of published data, exchanged and communicated via the web, new challenges for computer scientists arise. While the content within the internet-based communication can be encoded and decoded using a variety of types (in terms of media compilation and annotation), special methods and services need to be developed or adapted to be applied. Internet-based communication provides a challenging area to explore computational models of natural language texts and human communication on networked communication systems, promoting new research in computational linguistics and computer science.

Notes

1. http://www.viola.org/
2. http://mosaic.mcom.com/
3. http://browser.netscape.com/
4. http://altavista.digital.com
5. http://yahoo.com
6. http://www.google.de/
7. http://www.facebook.com/press/info.php?statistics
8. http://www.guardian.co.uk/technology/2011/jan/03/facebook-value-50bn-goldman-sachs-investment
9. http://www.cyc.com/opencyc/
10. http://www.linguistics-ontology.org/gold.html
11. http://www.wikimedia.org/
12. https://www.blogger.com/
13. https://www.twitter.com/
14. https://www.facebook.com/
15. http://delicious.com/
16. http://www.flickr.com/
17. The first text message was sent by Ray Tomlinson in 1971 via *ARPANET*.
18. https://wave.google.com/wave/

19. http://www.sipgate.de
20. http://wordpress.org/
21. A content management system is a system used to manage and edit content from a web-site without needing the expertise of the Hypertext Markup Language (HTML).
22. http://www.bibsonomy.org/
23. http://images.google.com/imagelabeler/
24. http://www.mediawiki.org/wiki/MediaWiki
25. http://www.facebook.com
26. http://www.twitter.com
27. http://www.myspace.com
28. http://www.linkedin.com
29. http://www.tumblr.com
30. http://www.xing.com/

References

Abbate, J.
2000 *Inventing the Internet*. MIT Press, Cambridge, Massachusetts.

Aery, Manu and Sharma Chakravarthy
2005 eMailSift: Email classification based on structure and content. In *ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 18–25, Washington, DC, USA. IEEE Computer Society.

Aitken, J. S., B. L. Webber and J. B. L. Bard
2004 Part-Of Relations in Anatomy Ontologies: A Proposal for RDFS and OWL Formalisations. *Pacific Symposium on Biocomputing*, 8:166–177.

Alby, Tom (ed.)
2008 *Web 2.0: Konzepte, Anwendungen, Technologien*. Carl Hanser, 3. edition.

Angevine, Duffy and A. Nur Zincir-Heywood
2008 A preliminary investigation of skype traffic classification using a minimalist feature set. In *Proceedings of the The Third International Conference on Availability, Reliability and Security, ARES 2008, March 4–7, 2008, Technical University of Catalonia, Barcelona*, pages 1075–1079.

Antoniou, Grigoris and Frank van Harmelen
2004 *A Semantic Web Primer*. MIT Press.

Armbrust, Michael, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica and Matei Zaharia
2009 Above the clouds: A berkeley view of cloud computing. Technical Report UCB/EECS-2009-28, Electrical Engineering and Computer Sciences, University of California at Berkeley. URL http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html.

Auer, Soren, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak and Zachary Ives
2007 DBpedia: a nucleus for a web of open data. In *Proceedings of the 6th International and the 2nd Asian Semantic Web Conference*, ISWC'07/ASWC'07, pages 722–735, Berlin, Heidelberg. Springer.

Back, Andrea, Norbert Gronau and Klaus Tochtermann
2008 *Web 2.0 in der Unternehmenspraxis: Grundlagen, Fallstudien und Trends zum Einsatz von Social Software*. Oldenbourg, München.
Backstrom, Lars, Eric Sun and Cameron Marlow
2010 Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 61–70, New York, NY, USA. ACM.
Baun, Christian, Marcel Kunze, Jens Nimis and Stefan Tai (eds.)
2011 *Cloud Computing – Web-basierte dynamische IT-Services*. Informatik im Fokus. Springer.
Beckett, Dave
2004 RDF/XML Syntax Specification (Revised). W3C Recommendation, W3C. URL http://www.w3.org/TR/rdf-syntax-grammar/.
Beißwenger, Michael
2007 *Sprachhandlungskoordination in der Chat-Kommunikation*, volume 26 of *Linguistik – Impulse & Tendenzen*. De Gruyter, Berlin/New York.
Beißwenger, Michael and Angelika Storrer
2010 Corpora of computer-mediated communication. In Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics. An International Handbook of the Science of Language and Society*, pages 292–308, Berlin/Boston. De Gruyter.
Berners-Lee, Tim
1994 Universal Resource Identifiers in WWW. Request for Comments 1630, Network Working Group. URL http://www.ietf.org/rfc/rfc1630.txt.
Berners-Lee, Tim
2011 Semantic web road map. http://www.w3.org/DesignIssues/Semantic.html, 1998. Retrieved on August 29.
Berners-Lee, Tim and Mark Fischetti
2000 *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper-Collins, San Francisco.
Berners-Lee, Tim, James Hendler and Ora Lassila
2001 The semantic web. *Scientific American Magazine*, 284(5):34–43.
Bhanu, S. Vijay, R. M. Chandrasekaran and V. Balakrishnan
2010 Effective bandwidth utilization in IEEE802.11 for VOIP. *International Journal of Computer Science and Information Security*, 8(1):68–75.
Bird, Daniel
2011 Teenager's remorse over facebook house party. http://www.telegraph.co.uk/technology/facebook/8561253/Teenagers-remorse-over-Facebook-house-party.html, 2011. Retrieved on August 28.
Birkenbihl, Klaus
2006 Standards für das Semantic Web. In Tassilo Pellegrini and Andreas Blumauer, editors, *Semantic Web – Wege zur vernetzten Wissensgesellschaft*, pages 73–88. Springer-Verlag.
Blood, Rebecca
2002 *The Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog*. Perseus Books, Cambridge MA.
Blumauer, Andreas and Tassilo Pellegrini (eds.)
2009 *Social Semantic Web: Web 2.0 – Was nun?* X.media.press. Springer, Berlin/New York.

Böhringer, Martin and Stuart J. Barnes
2009 Continuance usage intention in microblogging services: The case of twitter. In *Proceedings of the 17th European Conference on Information Systems (ECIS), Verona, Italy, June 08–10, 2009*, page in press.

Boyd, Danah
2008 *Taken out of Context. American Teen Sociality in Networked Publics.* PhD thesis, University of California. URL http://www.danah.org/papers/TakenOutOfContext.pdf.

Brants, Sabine, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius and George Smith
2002 The TIGER treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*, pages 24–41.

Brickley, Dan and R.V. Guha
2004 RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, W3C. URL http://www.w3.org/TR/rdf-schema/.

Brickley, Dan and Libby Miller
2010 FOAF vocabulary specification 0.98. http://xmlns.com/foaf/spec/, 2010. Retrieved on September 18.

Bush, Vannevar
1945 As we may think. *Atlantic Monthly*, 176(1):641–649.

Caldarelli, Guido
2007 *Scale-Free Networks: Complex Webs in Nature and Technology.* Oxford University Press, Oxford.

Campbell, Jeffrey D.
2004 Instant messages: a framework for reading between the lines. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 519–522, New York, NY, USA. ACM.

Canali, Claudia, Michele Colajanni and Riccardo Lancellotti
2011 Data acquisition in social networks: Issues and proposals. In *Proceedings of the International Workshop on Services and Open Sources (SOS'11), Brest, France*.

Carvalho, Vitor R. and William W. Cohen
2005 On the collective classification of email "speech acts". In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 345–352, New York, NY, USA. ACM.

Cattuto, Ciro, Christoph Schmitz, Andrea Baldassarri, Vito Servedio, Vittorio Loreto, Andreas Hotho, Miranda Grahl and Gerd Stumme
2007 Network properties of folksonomies. *AI Communications*, 20(4):245–262.

Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto and Krishna Gummadi
2010 Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

Chernov, Sergey, Tereza Iofciu, Wolfgang Nejdl and Xuan Zhou
2006 Extracting semantic relationships between Wikipedia categories. In Max Völkel and Sebastian Schaffert, editors, *Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics*, Workshop on Semantic Wikis. ESWC2006. URL http://semwiki.org/semwiki2006.

Chishti, M. Ahsan, Shaima Qureshi, A. H. Mir, Shariq Haseeb and Iftikhar Ahmad
2009 Utilizing semantic web as communication protocol in faded information field (FIF) architecture for information retrieval and dissemination. *The International Journal of Computing and ICT*, 3(2):41–46.

Damme, Céline Van, Martin Hepp and Katharina Siorpaes
2007 FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. In *Proceedings of the ESWC 2007 Workshop Bridging the Gap between Semantic Web and Web 2.0*, pages 71–84, Innsbruck, Austria.

de Kerckhove, Derrick, Martina Leeker and Kerstin Schmidt
2008 *McLuhan neu lesen: Kritische Analysen zu Medien und Kultur im 21. Jahrhundert*. Transcript, Bielefeld.

December, John
1996 Units of analysis for internet communication. *Journal of Computer-mediated Communication*, 1(4).

December, John
2010 Notes on defining of computer-mediated communication. http://www.december.com/cmc/mag/1997/jan/december.html, 1997. Retrieved on September 18.

Diewald, Niels
2012 Decentralized social network sites. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Ding, Ying, Ioan Toma, Sin-Jae Kang, Zhixiong Zhang and Michael Fried
2008 Mediating and analyzing social data. In *OTM '08: Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part II on On the Move to Meaningful Internet Systems*, pages 1355–1366, Berlin/Heidelberg. Springer.

Drezner, Daniel and Henry Farrell
2008 Introduction: Blogs, politics and power: a special issue of public choice. *Public Choice*, 134:1–13.

Dublin Core Metadata Initiative
2010 Dublin core metadata element set, version 1.1. Technical report, DCMI. URL url{http://dublincore.org/documents/2010/10/11/dces/}.

Erbach, Gregor
2004 Mapping, measuring, and modelling the diffusion of linguistic material on the internet. In *Proceedings of the First International Workshop on the Emergence and Evolution of Linguistic Communication*, Kanazawa.

Farrell, Henry and Daniel Drezner
2008 The power and politics of blogs. *Public Choice*, 134:15–30.

Gabrilovich, Evgeniy and Shaul Markovitch
2007 Computing semantic relatedness using Wikipedia-based explicit semantic analysis. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 6–12.

Garrett, Jesse James
2005 Ajax: A new approach to web applications. http://adaptivepath.com/ideas/essays/archives/000385.php.

Garton, Laura, Caroline Haythornthwaite and Barry Wellman
1997 Studying online social networks. *Journal of Computer-Mediated Communications*, 3 (1).

Gates, Bill
2000 Shaping the internet age. http://www.microsoft.com/presspass/exec/billg/writing/shapingtheinternet.mspx. Retrieved on September 18, 2010.

Glass, Richard and Suhong Li
2010 Social influence and instant messaging adoption. *Journal of Computer Information Systems*, W2010:24–30.

Gleim, Rüdiger, Alexander Mehler, Matthias Dehmer and Olga Pustylnikov
2007 Aisles through the category forest – utilising the Wikipedia category system for corpus building in machine learning. In *Proceedings of the 3rd International Conference on Web Information Systems and Technologies (WEBIST '07), March 3–6, 2007, Barcelona*, pages 142–149.

Golder, Scott A. and Bernardo A. Huberman
2006 Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198208.

Gruber, Thomas R.
1993 A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220.

Guha, R., Rob McCool and Eric Miller
2003 Semantic search. In *Proceedings of the 12th international conference on the World Wide Web (WWW '03)*, pages 700–709, New York, NY, USA. ACM.

Hammwöhner, Rainer
2007a Wikipedia – ein Medium der Ignoranz? In Achim Geisenhanslüke, editor, *Ignoranz: Nichtwissen, Vergessen und Missverstehen in Prozessen kultureller Transformationen*, volume 3 of *Literalität und Liminalität*, pages 229–258. transcript, Bielefeld.

Hammwöhner, Rainer
2007b Interlingual aspects of wikipedias quality. In *Proceedings of the 12th International Conference on Information Quality*, volume 1, page 3949.

Hancock, Jeffrey T., Lauren Curry, Saurabh Goorha and Michael Woodworth
2005 Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) – Track 1 – Volume 01*, HICSS '05, pages 22.3–, Washington, DC, USA. IEEE Computer Society.

Hauslohner, Abigail
2011 Is Egypt about to have a facebook revolution? http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html. Retrieved on August 28, 2011.

Hayati, Pedram and Vidyasagar Potdar
2008 Evaluation of spam detection and prevention frameworks for email and image spam: a state of art. In *Proceedings of the 10th International Conference on Information Integration and Web-based Applications and Services*, pages 520–527, New York, NY, USA. ACM.

Hayes, Patrick
2004 RDF semantics. http://www.w3.org/TR/2004/REC-rdf-mt-20040210/. Retrieved on September 18, 2010.

Hepp, Martin, Pieter De Leenheer, Aldo de Moor and York Sure (eds.)
2008 *Ontology Management, Semantic Web, Semantic Web Services, and Business Applications*, volume 7 of *Semantic Web And Beyond Computing for Human Experience*. Springer.

Herman, Ivan
2001 W3C semantic web activity. http://www.w3.org/2001/sw/. Retrieved on September 18, 2010.

Herring, Susan C., Inna Kouper, Lois Ann Scheidt and Elijah L. Wright
2004 Women and children last: The discursive construction of weblogs. In *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*, page 2004.

Hitzler, Pascal, Markus Krötzsch and Sebastian Rudolph
Foundations of Semantic Web Technologies. Chapman & Hall/CRC, 2009.

Philipp Hoschka
2011 Ubiquitous web domain. http://www.w3.org/UbiWeb/. Retrieved on August 29, 2011.

Hotho, Andreas, Robert Jäschke, Christoph Schmitz and Gerd Stumme
2006 Bibsonomy: A social bookmark and publication sharing system. In *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, pages 87–102, Aalborg. Aalborg University Press.

Howe, Jeff
2008 *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Business, 1. edition.

Isaacs, Ellen, Alan Walendowski, Steve Whittaker, Diane J. Schiano and Candace Kamm
2002 The character, functions, and styles of instant messaging in the workplace. In *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 11–20, New York, NY, USA. ACM.

Ito, Mizuko, Sonja Baumer and Matteo Bittanti
2009 *Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning with New Media*. MIT Press.

Java, Akshay, Xiaodan Song, Tim Finin and Belle Tseng
2007a Why we twitter: Understanding microblogging usage and communities. In *Procedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, College Park, Maryland. Computer Science and Electrical Engineering, University of Maryland.

Java, Akshay, Xiaodan Song, Tim Finin and Belle Tseng
2007b Why we twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pages 56–65.

Joisten, Martina
2007 Multimediale Gespräche in Skype: Hybridisierung von Gebrauchsweisen in der interpersonalen Kommunikation. In Simone Kimpeler, Michael Mangold, and Wolfgang Schweiger, editors, *Die digitale Herausforderung – Zehn Jahre Forschung zur computervermittelten Kommunikation*. Springer.

Jones, Steven G.
1998 Information, internet, and community in the information age: notes toward an understanding of community. In Steven G. Jones, editor, *Cybersociety 2.0*, pages 1–34. Sage Publications, Thousand Oaks, CA, USA.

Kalyanpur, Aditya, Evren Sirin, Bijan Parsia and James Hendler
2004 Hypermedia inspired ontology engineering environment: Swoop. In *Proc. International Semantic Web Conference (ISWC)*, pages 7–11.

Katz, James E. and Ronald E. Rice
2002 *Social Consequences of Internet Use: Access, Involvement, and Interaction*. The MIT Press, Cambridge.

Klyne, Graham and Jeremy Carroll
2004 Resource Description Framework (RDF): Concepts and abstract syntax. http://www.w3.org/TR/rdf-concepts/. Retrieved on September 18, 2010.

Knublauch, Holger, Ray W. Fergerson, Natalya F. Noy and Mark A. Musen
2004 The Protégé OWL plugin: An open development environment for semantic web applications. In *3rd International Semantic Web Conference (ISWC 2004)*, Hiroshima, Japan.

Kolari, Pranam, Tim Finin, Akshay Java, Anupam Joshi and Justin Martineau
2006 Blog track open task: Spam blog classification. In *TREC 2006 Blog Track Notebook*.

Kraut,R., M. Patterson, V. Lundmark, S. Kiesler, T. Mukopadhyay and W. Scherlis
1998 Internet paradox. A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53(9): 1017–1031.

Krötzsch, Markus, Denny Vrandecic and Max Völkel
2005 Wikipedia and the semantic web – the missing links. In *Proceedings of the WikiMania2005*.

Lassila, Ora and James A. Hendler
2007 Embracing "web 3.0". *IEEE Internet Computing*, 11(3):90–93.

Lee, Kyumin, James Caverlee and Steve Webb
2010 Uncovering social spammers: social honeypots + machine learning. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 435–442, New York, NY, USA. ACM. URL http://doi.acm.org/10.1145/1835449.1835522.

Lenhart, Amanda and Susannah Fox
2009 Twitter and status updating. Technical report, Pew Internet and American Life Project. URL http://www.pewinternet.org/Reports/2009/Twitter-and-status-updating.aspx.

Leuf, B. and W. Cunningham
2001 *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley, Boston, pap/cdr edition.

Manola, Frank and Eric Miller
2004 RDF Primer, W3C Recommendation. http://www.w3.org/TR/rdf-primer/. Retrieved on September 18, 2010.

May, Wolfgang
2006 Reasoning im und für das Semantic Web. In Tassilo Pellegrini and Andreas Blumauer, editors, *Semantic Web – Wege zur vernetzten Wissensgesellschaft*, pages 485–503. Springer-Verlag.

Mayer-Schonberger, Viktor
2009 *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press.

McGuinness, Deborah L. and Frank van Harmelen
2004 OWL Web Ontology Language overview. W3C recommendation, World Wide Web Consortium. URL url{http://www.w3.org/TR/owl-features}. Retrieved on September 18, 2010.

Mehler, Alexander
2006 Text linkage in the wiki medium – a comparative study. In Jussi Karlgren, editor, Proceedings of the EACL Workshop on New Text – Wikis and blogs and other dynamic text sources, April 3–7, 2006, Trento, Italy, pages 1–8.

Mehler, Alexander
2008 Structural similarities of complex networks: A computational model by example of wiki graphs. *Applied Artificial Intelligence*, 22 (7–8):619–683.

Mehler, Alexander
2011 A quantitative graph model of social ontologies by example of Wikipedia. In Matthias Dehmer, Frank Emmert-Streib, and Alexander Mehler, editors, *Towards an Information Theory of Complex Networks: Statistical Methods and Applications*. Birkhäuser, Boston/Basel.

Mehler, Alexander and Ulli Waltinger
2009 Enhancing document modeling by means of open topic models: Crossing the frontier of classification schemes in digital libraries by example of the DDC. *Library Hi Tech*, 27(4):520–539.

Melville, Prem, Wojciech Gryc and Richard D. Lawrence
2009 Sentiment analysis of blogs by combining lexical knowledge with text classification. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1275–1284, New York, NY, USA. ACM.

Mika, Peter
2004 Social networks and the semantic web. In *WI '04: Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 285–291, Washington, DC, USA. IEEE Computer Society.

Miller, Eric
2001 Digital libraries and the semantic web. http://www.w3.org/2001/09/06-ecdl/. Retrieved on September 18, 2010.

Mills, David
2009 What is the role of cloud computing, web 2.0, and web 3.0 semantic technologies in an era of connected governance? In *Proceedings of the Semantic Community Workshop and Web Conference on Semantic Exchange*, Fairfax, VA.

Milne, David and Ian H. Witten
2008 Learning to link with Wikipedia. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge mining*, pages 509–518, New York, NY, USA. ACM.

Milne, David, Olena Medelyan and Ian H. Witten
2006 Mining domain-specific thesauri from Wikipedia: A case study. In *ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*, pages 442–448.

Minkov, Einat, William W. Cohen and Andrew Y. Ng
2006 Contextual search and name disambiguation in email using graphs. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 27–34, New York, NY, USA. ACM.

Miura, Asako and Kiyomi Yamashita
2007 Psychological and social influences on blog writing: An online survey of blog authors in Japan. *Journal of Computer-Mediated Communication*, 12(4):1452–1471.

Mulvenna, Maurice D., Sarabjot S. Anand and Alex G. Büchner
2000 Personalization on the net using web mining: introduction. *Communication of the ACM*, 43(8):122–125.

Mustafaraj, Eni and Panagiotis Metaxas
2010 From obscurity to prominence in minutes: Political speech and real-time search. In *WebSci10: Extending the Frontiers of Society On-Line*.

Nakayama, Kotaro, Takahiro Hara and Shojiro Nishio
2008 Wikipedia link structure and text mining for semantic relation extraction. In Stephan Bloehdorn, Marko Grobelnik, Peter Mika, and Duc Thanh Tran, editors, *Proceedings of the Workshop on Semantic Search (SemSearch 2008) at the 5th European Semantic Web Conference (ESWC 2008)*, volume 334 of *CEUR Workshop Proceedings*. CEUR-WS.org.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge and Noah A. Smith
2010 From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

O'Reilly, Tim
2005 O'reilly – what is web 2.0. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html. Retrieved on September 18, 2010.

Paaß, Gerhard
2012 Document classification, information retrieval, text and web mining. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Park, Han Woo and Mike Thelwall
2003 Hyperlink analyses of the World Wide Web: A review. *Journal of Computer-mediated Communication*, 8(4).

Patel-Schneider, Peter, Patrick Hayes and Ian Horrocks
2001 DAML+OIL reference description. http://www.w3.org/TR/2001/NOTE-daml+oil-reference-20011218. Retrieved on September 18, 2010.

Patel-Schneider, Peter, Patrick Hayes and Ian Horrocks
2004 OWL Web Ontology Language semantics and abstract syntax. http://www.w3.org/TR/owl-semantics/. Retrieved on September 18, 2010.

Prud'hommeaux, Eric and Andy Seaborne
2008 SPARQL query language for RDF. http://www.w3.org/TR/rdf-sparql-query/. Retrieved on September 18, 2010.

Raymond, Eric S.
2001 *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary.* O'Reilly & Associates, Inc., Sebastopol, CA, USA.

Rose, T., M. Stevenson and M. Whitehead
2002 The Reuters Corpus Volume 1 – from yesterday's news to tomorrow's language resources. In *3rd International Conference on Language Resources and Evaluation*, pages 827–832.

Santini, Marina, Alexander Mehler and Serge Sharoff
2010 Riding the rough waves of genre on the web: Concepts and research questions. In Alexander Mehler, Serge Sharoff, and Marina Santini, editors, *Genres on the Web: Computational Models and Empirical Studies*, pages 3–32. Springer, Berlin/New York.

Sasaki, Felix and Arle Lommel
2012 Multilingual computing. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Schaffert, Sebastian, François Bry, Joachim Baumeister and Malte Kiesel
2008 Semantic wikis. *IEEE Software*, 25(4):8–11.

Schmidt, Jan
2009 *Das neue Netz – Merkmale, Praktiken und Folgen des Web 2.0*. UVK Verlagsgesellschaft.

Schuff, David, Ozgur Turetken, John D'Arcy and David Croson
2007 Managing e-mail overload: Solutions and future challenges. *Computer*, 40:31–36.

Sharoff, Serge and Anthony Hartley
2012 Lexicography, terminology and ontologies. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

SixApart
2002 Trackback technical specification. http://www.sixapart.com/pronet/docs/trackback _spec. Retrieved on September 18, 2010.

Spivack, Nova
2007 Radar networks towards a webos. http://novaspivack.typepad.com/RadarNetworksTowardsAWebOS.jpg. Retrieved on June 1, 2011.

Staab, Steffen and Rudi Studer (eds.)
2004 *Handbook on Ontologies*. International Handbooks on Information Systems. Springer, Berlin.

Staab, Steffen, Pedro Domingos, Peter Mika, Jennifer Golbeck, Li Ding, Tim Finin, Anupam Joshi, Andrzej Nowak and Robin R. Vallacher
2005 Social networks applied. *IEEE Intelligent Systems*, 20(1):80–93.

Stocker, Alexander and Klaus Tochtermann
2009 Anwendungen und Technologien des Web 2.0: Ein Überblick. In Andreas Blumauer and Tassilo Pellegrini, editors, *Social Semantic Web: Web 2.0 – Was nun?*, X.media.press, pages 63–82. Springer, Berlin/New York.

Storrer, Angelika
2001 Sprachliche Besonderheiten getippter Gespräche. Sprecherwechsel und sprachliches Zeigen in der Chat-Kommunikation. In Michael Beißwenger, editor, *Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in*

synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld, pages 3–24. Ibidem Verlag, Hannover.

Strube, M. and S. Paolo-Ponzetto
2006 WikiRelate! Computing semantic relatedness using Wikipedia. In *Proceedings of the 21st National Conference on Artificial Intelligence*, Boston, USA.

Stuckenschmidt, Heiner and Frank van Harmelen
2005 *Information Sharing on the Semantic Web*. Springer.

Stührenberg, Maik
2012 Foundations of markup languages. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Stumme, Gerd, Andreas Hotho and Bettina Berendt
2011 Semantic web mining. State of the art and future directions. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(2):124–143.

Sun, Aixin, Maggy Anastasia Suryanto and Ying Liu
2007 Blog classification using tags: an empirical study. In *ICADL'07: Proceedings of the 10th international conference on Asian digital libraries*, pages 307–316, Berlin, Heidelberg. Springer.

Surowiecki, James
2004 *The Wisdom of Crowds*. Random House, New York.

Tapio, Antti
2005 Future of telecommunication – internet telephone operator skype. HUT T-110.551 Seminar on Internetworking.

Thurlow, Crispin, Laura Lengel and Alice Tomic
2004 *Computer Mediated Communication*. Sage Publications, London.

Trippel, Thorsten
2012 Controlled language structures in technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Vakali, A. and G. Pallis
2003 Content delivery networks: status and trends. *Internet Computing, IEEE*, 7(6):68–74.

Vivacqua, Adriana S. and Marcos R.S. Borges
2010 Collective intelligence for the design of emergency response. In *Proceedings of the 2010 14th International Conference on Computer Supported Cooperative Work in Design*, pages 623–628.

Voss, Jakob
2005 Measuring Wikipedia. In *International Conference of the International Society for Scientometrics and Informetrics: 10th*. ISSI.

W3C
2009 W3C semantic web activity. http://www.w3.org/2001/sw/. Retrieved on September 18, 2010.

Wal, Thomas Vander
2005 Folksonomy. http://www.vanderwal.net/essays/051130/folksonomy.pdf. Retrieved on August 9, 2011.

Waltinger, Ulli and Alexander Mehler
2009 Social semantics and its evaluation by means of semantic relatedness and open topic models. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology – Volume 01*, WI-IAT '09, pages 42–49, Washington, DC, USA. IEEE Computer Society.

Waltinger, Ulli, Irene Cramer and Tonio Wandmacher
2009 From social networks to distributional properties: A comparative study on computing semantic relatedness. In *Proceedings of the Annual Meeting of the Cognitive Science Society – CogSci 2009, Amsterdam (NL), 2009*.

Walton, Christopher D.
2006 *Agency and the Semantic Web*. Oxford University Press.

Wilson, Christo, Bryce Boe, Alessandra Sala, Krishna P.N. Puttaswamy and Ben Y. Zhao
2009 User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, EuroSys '09, pages 205–218, New York, NY, USA. ACM. URL http://doi.acm.org/10.1145/1519065.1519089.

Xiang, Yang, Wanlei Zhou and Jinjun Chen
2007 Managing email overload with an automatic nonparametric clustering approach. In *Proceedings of IFIP International Conference on Network and Parallel Computing (NPC 2007)*, pages 81–90.

Youn, S. and D. McLeod
2007 A comparative study for email classification. In Khaled Elleithy, editor, *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, pages 387–391. Springer, Dordrecht.

Zesch, Torsten, Christof Müller and Iryna Gurevych
2008 Extracting lexical semantic knowledge from Wikipedia and Wiktionary. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco.