The availability, processing, and usage of any kind of information, as well as the way in which we communicate, has changed dramatically during the last 15 years, especially through the technologies associated with the Internet. Today, a vast amount of human knowledge is available at the click of a button. New services will widen the Internet into an interactive medium for social networks (Web 2.0) and interconnect the physical world (cyber-physical systems) in an unprecedented manner. Usage of this future Internet will continue to evolve rapidly during the next decade and beyond. This development will result in many new challenges from different fields of research, and therefore have to be investigated from different angles. Under the motto ‘Internet, Information and I – Living in a Connected World’, the aim of this interdisciplinary research at L3S and the Leibniz Universität Hannover is to investigate the growing connections between human individuals and the increasing amount of information. This work will examine the impact of machines and persons in virtual, real and social networks. The research initiative is divided into four workgroups.

Internet of Things

The term “Internet of Things” is associated with wireless, self-configuring networks of real-world objects. If all devices of our daily lives, i.e. everything from cars and small devices up to cars are connected in this way (for example with RFID-tags), they can be identified, managed, tracked and used by computers. Naturally, this raises a vast number of previously untackled problems. However, there are a multitude of new possibilities for effectively using this new data, presenting enormous potential for innovative products and new markets.

Audio-Visual Data

The aim of this group is to focus on challenges surrounding the accessibility of audio-visual data. The usage of audio, video, and pictures is an important part of today’s information society. New technologies open up the possibilities of utilizing them independent of time (for example, in time- shift broadcasts) and place (at the working place, or in a car or train). In spite of the wide use and availability of new technologies, they are rarely viable for the end user, because of their complexity. Storage, effective search, and efficient distribution are also major fields of research here.

Privacy & Copyright

For all these new technologies, the aspects of data privacy and data security, as well as intellectual property rights are important. The implications of these new ideas will be examined and made clear in the early stages of the development of projects, to ensure compliance with existing laws and standards. Additionally, privacy issues – from both a scientific and political point of view – will be addressed. Personalized applications are one of the most important perspectives of the future Internet. A broad legal evaluation and development of strategies for the implementation of privacy will be performed inside this group.

Services & Architectures

An infrastructure with high performance is crucial due to the extensive number of services that are involved. This includes high processing power, as well as fast transmission and a high data rate for any type of data. The architecture, topology, and bandwidth has to scale along with the dramatically-changing requirements; such requirements include grid and cloud computing services.

Web Science – Investigating the Future of Information and Communication

Information production and consumption has dramatically changed during the last 20 years. While information was primarily distributed in printed form during the last 500 years, information distribution is now rapidly changing to electronic form. With the explosive advent of the World Wide Web, more than 2 billion users, all over the world, can now access more than 40 billion static pages over the Internet and an almost infinite amount of dynamic content. Today, most of the world’s knowledge is available through the Web and its interconnected databases and digital libraries. Furthermore, the Web has become a highly interactive medium, where users can not only consume, but also produce new information. Yet, due to its unprecedented growth and the wealth of supported applications, the Web poses more challenges than 16 years ago. These challenges can only be solved with multi-disciplinary methods and technologies from information systems and computer science, with additional input from social science, business and law. The L3S Research Center is at the forefront of this new field of Web Science, in cooperation with partners world-wide, developing innovative methods and technologies to support and connect individuals and communities in all sectors of the knowledge society, and harnessing the information ecosystem to connect today’s Web to real-world entities.

Several interrelated topics guide our research in this context: The Web of People provides the focal point of our efforts, acknowledging the central role people play in the complex ecosystem of the Web, including work on social networks, human computing and trust and reputation. Our research into Foundations of Web Science and Web Infrastructure provide the algorithmic and conceptual foundations for the Web as well as the necessary infra-structures to ubiquitously access information everywhere and at all times, and to connect the Web with the real world. Web Information Management and Web Search play a central role for how to access information on the Web, in the context of Digital Libraries and the Deep Web, as well as for accessing all kinds of unstructured, semi-structured and multimedia information on the Web. Last, but not least, service computing as well as advanced security and privacy mechanisms provide important Software Technologies for Web Applications, for example in E-Science and E-Learning applications.
Interview with Ralf Steinmetz – Professor of Multimedia Communications at the Darmstadt University of Technology

The Future Internet – Challenges and Chances

Prof. Ralf Steinmetz worked for over nine years in industrial research and development of distributed multimedia systems and applications. He has been head, since 1996, of the Multimedia Communications lab at Darmstadt University of Technology, Germany.

From 1997 to 2001 Professor Dr. Ralf Steinmetz directed the Fraunhofer (former GMD) Integrated Publishing Systems Institute IPSI in Darmstadt. In 1999 he founded the Hessian Telemedia Technology Competence Center (httc e.V.). His thematic focus in research and teaching is on multimedia communications with his vision of real “seamless multimedia communications”. With over 200 refereed publications he has become ICCC Governor in 1999, was awarded the ranking of Fellow of both, the IEEE in 1999 and the ACM in 2002. Professor Dr. Ralf Steinmetz is a member of the Scientific Council and president of the Board of Trustees of the International research institute IMDEA Networks.

L3S: Professor Steinmetz, what do you think are the main challenges in setting up a Future Internet? Which challenges do you see on the network layer and which ones do we face on the application and content layer?

In the past, the Internet has been tremendously successful in connecting people and comput- ers. Recently, we have witnessed that distributed systems, such as peer-to-peer and user-generated content, have been driving forces behind the development of the Internet. Still, we do not yet have fully seamless communications, this is a main challenge that persists. For the users it is quite important, that the Internet can be used easily, reliably and securely; but also we need to consider the Quality of Service of certain classes of applications.

Within the Future Internet, we might have to depart radically from a layered architecture (i.e., there might not be a network layer or application layer). However, I foresee that the semantics of the delivered content or the context of commu- nications are crucial to control and enhance the experience delivered by the Internet – indepen- dently of whether it has layers or not. Last but not least, the design of the future Internet has to allow for further evolution. Imagine the inclusion of billions of sensor nodes into the “Internet of Things”, using today’s technologies we fall short in adequately including these devices seamlessly into the Internet due to various technical limita- tions.

L3S: What about potential privacy issues, when accessing the Internet everywhere and with any kind of devices? Do you see convincing solutions for privacy protection on the Future Internet?

Privacy and security are important challenges for a ubiquitous and seamless Internet. The main point I want to make here is that a technical system as complex as the Internet, can never be fully secure or privacy preserving. There is always a fight between the “good guys” and the “bad guys”. And I think that information that is related to the privacy of a user is well protected using state-of-the-art mechanisms as at today, in perhaps ten years, these mechanisms might be broken. We do need technical solutions to address privacy, moreover, we need to make the users aware, how secure (or insecure) the navigation and their information is. It should also be very clear that a technical solution to these problems is only part of the solution: society needs to discuss the implications of a digital world with respect to privacy and security.

L3S: The predominant business model of the Web is advertisement. Do you think there are other ideas around that are about to break through? What other business models do you consider valuable for the Future Internet?

Really? Is the predominant business model “advertisement”. I am not sure about this! The In- ternet allows access to information and services. Depending on what services or information we use and in which context, advertisement might be an appropriate business model as we witness today. But I want to be very critical: to get something for free it is certainly at- tractorive. To get it in a better quality and to pay for this is also possible. Small payments like the applications for an iPhone or for e.g. 0,79 EUR are also something we accept easily, because we perceive an added value. It is also clear that there are huge application areas such as E-commerce that rely on the Internet for critical business transactions and use business models other than advertisement.

L3S: One of the promises of the Future Internet is ubiquitous availability of information and media. Isn’t this going along with challenges in usability? How do you think will applications of the Future Internet manage to be usable by common users?

The ubiquitous Internet will only be successful (on the long run) if it is accepted widely by the users, i.e., it has to be easy to use and to be open to everyone. Young and old, poor and rich, everybody has to make use of it and take advantage of it. The digital divide is yet to be overcome. This presents a huge opportunity for innovators: bring services to users in an easy to use fashion; that is, add value for the “common user”, and you might be very successful.

L3S: Future Internet will probably include the Internet of Things. What kind of things do you believe are most interesting as part of that? Which scenarios do you think of in that context?

As mentioned earlier, the Internet of Things, with potentially billions of networked sensors, presents a huge technological challenge. Today, we see the first incarnations of such an Internet of Things. Logistics and supply chains are net- worked and goods are traded using sensors. This enables real-time information on the location of such goods, but also about the current state such as temperature, etc. In turn, this informa- tion allows for the real-time management of business processes. In the future, with sensors being ubiquitous (think of your cellphone as a sensor) we will step into an Internet that closes the gap between a virtual world and a real world by providing access to the context of objects in the real world and/or by manipulating and interacting with these objects.

L3S: Professor Steinmetz, we thank you very much for this interview.

LivingKnowledge
Fact, Opinions and Bias in Time

The Web is characterized by the diversity and multitude of its content. Cultural backgrounds, schools of thought, geographical contexts etc. contribute to this diversity. In this way, diverging viewpoints and conflicts are expressed in the Web. Time and evolution represent a further dimension making diversity an intrin- sic and unavoidable property of knowledge. Today’s search technologies do not provide means for dealing with this variety. Therefore, it is the goal of the LivingKnowledge project to develop innovative methods for informa- tion access and management that is aware of bias, diversity and evolution in information sources.

L3S: What is the role of the Future Internet in adequately including these devices seamlessly into the Internet due to various technical limita- tions?

The Web as source of information reflects the diversity and conflict regarding future events and statements, relying upon information already available on the Web. Based on the European Web Archive and coupled with Yahoo! News, it will constantly update the statements and predictions about the future, and make them available for searching and mining the future as part of the Yahoo!com Web portal. The Media Research Analyst will focus on opinion and bias, and address questions about the public image of a company, possibly changing over time, or the effectiveness of a PR campaign as reflected through user generated content in blogs and other public forums.

The leading search engine company Yahoo! and information providers such as the European Ar- chive have joined LivingKnowledge, because they understand the added value that diversity, evolution- and bias-aware search technologies can provide to many of their customers.

URL: http://livingknowledge-project.eu
Dr. Kerstin Denecke, denecke@L3S.de
Dr. Kerstin Denecke, denecke@L3S.de
Can All Tags Be Used for Search?

L3S researchers have improved the state-of-the-art in the area of collaborative tagging systems focusing on tagging motivations, tagging patterns and how tags can be used to improve search applications. Web2.0 tools and environments have made collaborative tagging very popular, people assign freely selected words, in the form of keywords or category labels, to shared content – thereby describing and organizing these resources. Given the rapid increase in popularity of Web2.0 tools in the last years, a huge amount of user generated metadata has become available.

However, users’ motivations for tagging resources, as well as the types of tags differ across systems. Typically, tags represent many different aspects of the resources they describe and it is not obvious whether and how these tags or subsets of them can be used for search. Some studies have started to investigate tagging motivations and patterns, but they are usually focusing on only one specific collection. Other work studies how to support the tagging process and focus on how to improve general information retrieval algorithms using tags. L3S researchers performed the first in-depth study of these aspects for different kinds of resources and systems – Web pages (Del.icio.us), music (Last.fm) and images (Flickr), thus advancing the state-of-the-art in this domain.

Our analysis revealed the necessity and usefulness of a common tag classification scheme for different collections, which allows the comparison of the types of tags used in different tagging environments. For example, the distributions of tag types strongly depend on the resources they annotate: for Flickr and Del.icio.us the most frequently used tags (50% of the cases) refer to topic concepts (i.e. what the resource is about), while for Last.fm, type-related tags (e.g. genres) are the most prominent ones. Other interesting results of our analysis refer to the added value of tags to existing content. More than 50% of existing tags bring new information to the resources they annotate. For the music domain this is even the case for 95.8% of the tags. A large amount of tags is accurate and reliable, in the music domain for example 73.01% of the tags also occur in online music reviews.

Caulking the Firewall Individually Authorized Access

Frequently, news appears in the media about computer viruses and trojans that infected personal computers and servers. Unauthorized access to personal or sensitive information is the result. Firewalls, which should restrict the access to these computers and protect the data, are more often unable to fulfill their mission due to unpredictable connection establishment and the use of encryption. To individually authorize connections to computers during the connection establishment phase, we developed a technique which enables access control to firewalls based on the user’s proven identity.

The spectacular news about the espionage and remote control of computers of the Office of His Holiness the Dalai Lama (OHHDL) is an example for the potential of intruders, which threaten sensitive information. But not only sensitive information from national and international organizations is a desirable object for attackers. Also credit card information and other financial data from common users are in the focus of attackers, which use viruses and trojans to access access to resources. These threats concern both personal computers and servers, which rely on standard operating systems and use standard software components; also, cutting edge developments and concepts are vulnerable to unauthorized access. Grids environments – with massive amount of computing and storage resources, and the value of the stored information are especially desirable targets for attackers.

Driven by the necessity to secure our resources, we developed a technique called TCP-AuthN, which provides a means to restrict the access to authorized users that are authenticated by their personal X.509 certificate. While a connection between a users computer (client) to an application providing server is initiated, a user’s identity is provided to firewalls on the path between the client and the server. These firewalls compare the presented information to a preconfigured list of users, affiliations or attributes and judge over the connection establishment. To establish a reliable connection between two computers over the Internet the Transport Control Protocol (TCP) is used. In the connection establishment phase, TCP exchanges three segments to synchronize internal counters. We use the first segment of this three-way handshake to transport information about where to retrieve the user’s certificate. The second and third segments of the TCP three-way handshake are used to accomplish a challenge response procedure to prevent man-in-the-middle and replay attacks, where attackers collect and replay segments of a previously established connection. Since it is possible to transport application data within these first three segments, we have to introduce a new TCP option, which enables the distinction of application data from data used for authentication purposes.

This new method TCP-AuthN, which enables the dynamic operation of firewalls based on the users verified identity, his affiliations or attributes within this affiliation enhances the security of firewalled personal computers and servers. Furthermore, the restriction is based on an individual means; therefore a fine-grained access restriction is possible.

Social networks are an important driver in today’s net-economy. They are successful because it’s convenient to spread and get information about every user online under the principle of making information easy to share.

This – by design – is in conflict with some of the basic principles of Europe’s legislation on Data Protection, that is: “personal data shall be processed fairly and lawfully, not kept for longer than necessary, shall be adequate, relevant and not excessive in relation to the purposes for which they were collected, accurate and up to date”. Most of today’s Web 2.0 services are available at no charge and are financed through targeted behavioral advertising – which in turn is spurring the collecting of personal data. Economic incentives to collect and process personal data therefore constantly compete with legal restrictions not to do so.

This is what happened in the past and is happening today, so – one could argue – why still bother about it? In a world of user-generated content, an increasing amount of personal data is available for everyone about everyone, without any restriction and time-limit. Examples of potentially sensitive information include: a posting in a forum on drug abuse a party-photo showing the person intoxicated, videos from a holiday trip, blog entries about personal issues, etc. Such content may seriously influence professional and personal affairs and there is no way to reverse the secrecy, once user content becomes publicly available on the net.

Of course, the individual has (in theory) inter alia, the right to access the data and to object on legitimate grounds against the processing of his or her data. But in practice, many problems emerge. The applicable law (thus affecting the user’s rights) is unclear: enforceability and liability rely on national rules and therefore differ depending on where the data is stored and processed. Furthermore legal claims are costly and complicated.

Regarding search, our studies showed that most of the tags can be used for search and that in most cases tagging behavior exhibits approximately the same characteristics as searching behavior. However, some noteworthy differences have also been observed. Namely, for the music domain, the usage context (i.e. situation suitable for listening to a particular song – e.g. “pool party”) is very useful for search, yet underrepresented in the tagging material. Similar, for pictures and music opinions or qualities (e.g. characteristics, moods) queries occur quite often, although people tend to neglect this category for tagging. Clearly, supporting and motivating tags within these categories could provide additional information valuable for search.

These results provide more insight into the use of different kinds of tags for improving search and they also indicate potential extensions of searching. It is also true that data protection regulations are difficult to read and might hinder developments that are technically possible. Finally, common standards on the minimum level of protection don’t exist yet on a global level.

These issues have not hampered rapid growth. The widespread collection and use of personal data without the subject’s consent will, if not treated properly, produce a digital tsunami in the near future: everybody might know everything about everybody. Consent is no way out and even if consent on the processing of personal data exists, its legality can often be challenged since it has to be a freely given, specific and informed. Revealing a once given consent, might not have any effect on the net which is – of course – not in line with the legal idea of consent and its revocation.

Data protection law will therefore have to deal with three main challenges: First, considering issues on a global level, in a world without global laws. Second, creating a legal network that has impact on the design of the web, without destroying its fascinating technical and social possibilities. Third, providing users with instruments needed to handle the most fundamental rights: to freely develop their identity.

Dealing with these challenges will be the main responsibility of the IRI (Institute for Legal Informatics) within the future of the internet initiative. IRI’s main goal will be to help the initiative to set a standard of good legal practice in the field and to become a driver for the entire consortium by playing the role of user advocate within the initiative.


erichard@iri.uni-hannover.de

URL: http://www.pharos-audiovisual-search.eu

Contact: Rebecca Pen, pharos@iri.uni-hannover.de

URL: http://www.rrzn.uni-hannover.de/tcpauthn.html

Contact: Prof. Dr. Nikolaus Forgh, nikolaus.forgo@uni-hannover.de

URL: http://www.irri.uni-hannover.de/lcai.html

Contact: Jan Winkelhöfer, winkelhofer@iri.uni-hannover.de

URL: http://www.zcc.uni-hannover.de/lcai.html

Contact: Jen Winkelhöfer, winkelhofer@iri.uni-hannover.de

URL: http://www.pharis.uni-hannover.de/search.ev
Detecting Attractive Photos

L3S researchers have developed a novel method for automatically determining the attractiveness of photos. The work is based on two aspects: (1) image features known for having significant effects on the visual quality perceived by humans (e.g. sharpness and colorfulness) and (2) community feedback and text annotations available in the Web 2.0 photo sharing system Flickr. Combining this information and applying techniques from the field of machine learning, the researchers developed a system for automatically estimating the attractiveness of photos.

Image attractiveness is a very subjective concept influenced by a wide number of factors. Even though artistic quality cannot be quantitatively computed, it correlates to certain visual features of images, assigning more optimal values to them. For instance, appealing images tend to have higher colorfulness, increased contrast and sharpness. In the social photo sharing environment, Flickr photos are accompanied by a variety of annotations such as tags, number of views, user comments, upload date, etc. For instance, the number of views is an indicator for the popularity of a photo. The upload date and the date a photo was taken are indicators for the recency of the content. In addition, adding a photo to one’s favourite list is probably the most direct positive relevance assignment in Flickr, and is an explicit expression of interest in the photo.

For recently uploaded photos or private collections, community feedback in any form might not (yet) be available. L3S researchers have used favorite assignments in Flickr to obtain statistical information for a community-based notion of "attractiveness" from a large number of photos and their connection to the mentioned visual features and tag annotations. Using methods from the area of machine learning the researchers have build models and software which is able to estimate the attractiveness of new photos for which no human feedback exists.

Legal Information:
Dipl.-Vw. Gabriele Herrmann-Krotz (Managing Director)
L3S Research Center
Appelstr. 9a
30167 Hannover, Germany
Tel.: +49 (0)511 762 - 17779
Fax: +49 (0)511 762 - 17779
E-Mail: herrmann@L3S.de
Internet: http://www.L3S.de

URL: http://creativecommons.org/licenses/by/3.0/de/deed.en