Future Directions of Knowledge Systems Environments for Web 3.0

Koji ZETTSU [a,1], Bernhard THALHEIM [b,2], Yutaka KIDAWARA [a,3], Elina KARTTUNEN [c,4], and Hannu JAAKKOLA [c,5]

[a] *National Institute of Information and Communications Technology, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289 Japan*

[b] *Christian-Albrechts-University Kiel, Computer Science Institute, 24098 Kiel, Germany*

[c] *Tampere University of Technology, P.O.Box 300, FI-28101 Pori, Finland*

Abstract. The internet and web applications have changed business and human life. Nowadays everybody is used to obtain data through the internet. Most applications are still Web 1.0 applications. Web 2.0 community collaboration and annotated data on the basis of Web 3.0 technologies supports new businesses and applications. The quality dimension of the web is however one of the main challenges. Knowledge systems target at high-quality data on safe grounds, with a good reference to established science and technology and with data adaptation to user's needs and demands. Knowledge system can be build based on existing and novel technologies. This paper discusses the challenges, two solutions and the fundamentals of knowledge system environments.

Keywords. web, information systems; knowledge web, next generation web; context, content, concept, and topic modelling; cloud computing, cloud services; knowledge system environments; information credibility, universal communication;

1. Introduction

1.1. Web x.0 Evolution and the Knowledge Web

For almost two decades the internet was a linkage of networked servers, which was entirely used as a worldwide source for researches. It resulted in an aggregate of billions of static web sites, which was accessed via hyperlinks. Websites have mainly been author-driven. They have been aiming to support users depending on their information need and demand, so the focus was chiefly on the mutual trust between user and provider. The utilisation of these sites can be modelled by

[1]zettsu@nict.go.jp
[2]thalheim@is.informatik.uni-kiel.de
[3]kidawara@nict.go.jp
[4]elina.karttunen@tut.fi
[5]hannu.jaakkola@tut.fi

story spaces. The story space specification results in storyboards that are schemes for utilisation by a large variety of users. Web 1.0 is author driven and uses as stories

- at the provider side `publish/provide_story/support` or `advertise/wait/attract/react/retain` and
- at the user side `inform/subscribe/obtain/answer/come_back`.

Web 1.0 has mainly be oriented towards content provision, which basically meant to deliver content together with a rudimentary functionality. These main functionalities can be:

- navigation facilities for inside site or page navigation;
- acquisition possibilities of information for users from simple content that is based on text, media data such as pictures, audio and video data;
- linking facilities;
- search and browse facilities providing to users.

Websites are mainly oriented towards content delivery, provide some functionality and are using a large variety of presentation facilities.

Web 1.0 has made author-driven static content available to numerous users. Users could access exclusively the web pages for researches and personal investigations. The control and management from the 'top' didn't provide any scope or client-side opportunities for development. This has changed with the evolution of Web 2.0, the so-called social web, as a development process powered by collaborative brainstorming, in which the collective cooperation is to the fore. Meanwhile there are no bounds set to the today's web. With the establishment of user communities, users obtain an abundance of information by high-tech sophisticated services, interchange experiences and benefits by the mass collaboration every single day, because data acquisition and data diffusion are basically accomplished by user interactions inside the whole web story space.

While Web 2.0 integrates collaboration, Web 3.0 provides annotation techniques. These annotation techniques are typically based on linguistic semantics of words used for a reference of data chunks to user semantics. These techniques provide a very good background for sophisticated search and representation techniques. Fully-developed Web 3.0 is characterised by the formula (4C + P + VS) where

- 4C means content, commerce, community and context
- P is used for personalization and
- VS denotes vertical search.

But what is missed in the future of web, is quality. We want to reach this level of quality with the aid of semantics and pragmatics in respect of the user profile and life cases. We are convinced that lexical semantics composes the base frame of the *Next Generation Knowledge Web*.

Figures 1 illustrates the general facets of websites. We distinguish six different facets: presentation (layout and playout) of pages within a website, (aggregated and prepared) data and functionality provided by the systems that support the website, stories and context behind the application logic of the website, and the user space that is based on a description of the intentions. Web 1.0 was mainly based on presentation systems with supporting systems for aggregated data (called content) and functionality. Web 2.0 allows context injection and is user-centered and story-centered. Web 3.0 extends the data content by annotation that are meaningful to users, i.e. provides content together with topical data. The knowledge web extends this dimension by explicit support for concepts beside annotations. It additionally allows an adaptation to the user and the context thus providing information the user really needs.

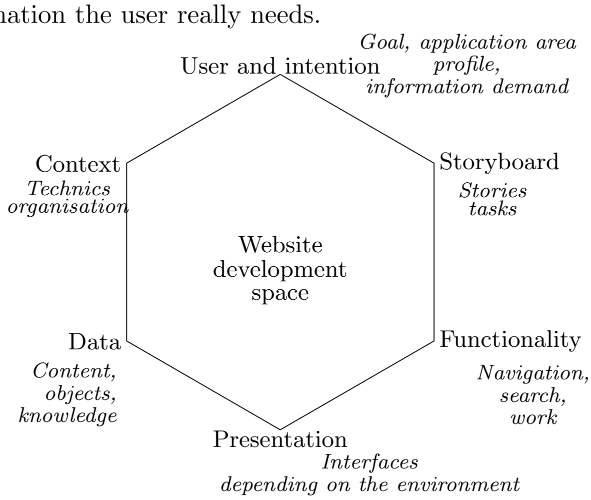

Figure 1. Separation of concern for development of Web x.0 websites

1.2. Knowledge Web - Do We Have a Need for That?

Human often meet a situation in which additional information, knowledge or at least fact are urgently demanded. This knowledge on demand is however not uniquely determined. It depends on the user, the current user situation, the data on hand, the background, the policies of data providers, etc.

`Example` (knowledge demand): *Let us consider the large variety for knowledge demand of people after the Iceland EyJafjallajokull Glacier volcano eruption on March 20, 2010:*

- *How long this situation will influence travel in Europe? Remember that the last Eyjafjallajokull eruption lasted for two years, and it is possible that this one will do the same. How weather conditions such as the anti-cyclone situation influence on ash spread?*
- *What are the contents of ash? Could particles of rock, glass and sand clog up aircraft engines? What are the fears of the effect of volcanic ash on plane engines? Are there other components on aircraft that are equally sensitive to particles? Is driving more dangerous than flying through ash? As flights resume, how dangerous is it to fly through a volcanic ash cloud? Are the airlines right with their requirement to resume flights on manual control by pilots depending on visibility? Which safety tests showed that the engines could cope in areas of low-density ash?*
- *Why mathematical simulations have been used for decision making? Why mathematics has partially failed in making predictions?*

- *How the weather changes can be explained after the volcano eruption? Why scientist were incorrect in their prediction for the weather impact? (The European summer in 2010 was far colder than any prediction could foresee. This summer seems to be a counterexample for the climate change discussion. Watching the enormous plumes of dust and ash rising from Eyjafjallajokull, it is hard to imagine that this almost week-long eruption would not have any effect on weather and climate. But scientist expected that there is no change.)*
- *What is the economical impact of such eruptions in general and of this eruption in special? What is the impact of the eruption for North Sea fishery, for industry, for tourism, etc.?*
- *What are the passengers rights for stranded passengers or cancellations? What are the best sources of advice? How I can cope with my personal situation? E.g., who gets priority on seats now flights are running again?*
- *Why icelanders enjoy their volcanos?*
- *How clouds depends on volcanos and flights? Jet contrails are effectively acting as cirrus clouds, reflecting solar energy in the day, acting as a blanket by night.*
- *Is there any correlation to other climate change drivers such as sun activity? What are the implications of ionospheric plasma bubbles? To what extent are sunspot activities related to economic cycles?*

This small list can be extended in many directions and illustrates the variety of knowledge that is necessary to satisfy the demand of people.

The example shows that we need different data, concepts, explanations, theories, and information. In general, knowledge system environments must support the following kinds:

- state-of-the art, -affairs, -knowledge, -science;
- deficiencies, missing or withhold facts;
- background, scientific explanations, science, potential theories, analysis;
- cross links, bindings;
- associations;
- facts with quality properties, full or partial picture;
- predictions, possible tactics and strategies for the future;
- restrictions, generalisation;
- analogies;
- history beside news;
- ways to cope with and the outcome for the future;
- consequences;
- links with headlines and quality assessment.

This list of knowledge pieces or chunks that must be provided can be categorized by the utility that the knowledge provides as follows:

Orientation knowledge allows to cope with the situation, to explain, and to survey the history, the scenario, the facts, the summarisation or generalisation and the overall view.

Tacit or action knowledge is based on practices, technics, methods, and strategies. It provides rules, procedures, check lists, principles, strategies, law, regulations, comments to regulations in order to manage situations.

Explanation knowledge gives reasons, arguments for explanation of claims or arguments or assertions or recommendations (what, why,, ...).

Sources knowledge links to knowledge on data sources (meta knowledge) such as knowledge on archives, references to communication, or cross links.

Activity knowledge supports working, adaptation or processing, operating on analogies, and coping with errors.

1.3. The Notion of Knowledge

The notion of knowledge[6] is one of overused terms. Knowledge has two sides. It is knowledge in general defined by a noun from one side and the knowledge by a user expressed by the verb 'to know' from the other side.

Knowledge as sustainable, potentially durable and verifiable grounded consensus: The required information can be qualified as knowledge, if the information

1. is consensus with a world and a community,
2. is based on postulates or principles that create the fundament for the knowledge,
3. is true according to a certain notion of 'truth',
4. is reusable in a rule system for new information,
5. is long-lasting and existing for a long time,
6. has an effect and is sustaining within a society, community or world, and
7. is not equivalent to other information that can be generated with the aid of facts or preliminary information in the particular inventory of knowledge by a rule system.

Knowledge as the state of information of a user: Different kinds of `'to know'` are:

1. The state or fact of knowing.
2. Familiarity, awareness, or understanding gained through experience or study.
3. The sum or range of what has been perceived, discovered or learned.

---

[6]The definition provided by the Encyclopedia Britannica [27] considers two 'Janus' meanings beside the obsolete 'cognizance' and the archaic 'sexual intercourse':
(I) as the fact of knowing something:
(Ia1) the fact or condition of knowing something with familiarity gained through experience or association;
(Ia2) acquaintance with or understanding of a science, art, or technique;
(Ib1) the fact or condition of being aware of something;
(Ib2) the range of one's information or understanding;
(Ic) the circumstance or condition of apprehending truth or fact through reasoning or cognition;
(Id) the fact or condition of having information or of being learned;
(II) the body of things known about or in science:
(IIa) the sum of what is known: the body of truth, information, and principles acquired by mankind;
(IIb) a branch of learning (synonyms of knowledge: learning, erudition, scholarship) meaning what is or can be known by an individual or by mankind.
We prefer this approach over the approach taken by the `Wikipedia` community who distinguishes between communicating knowledge, situated knowledge, partial knowledge, scientific knowledge and know-how or know-what or know-why or know-who knowledge.

4. Learning; erudition: teachers of great knowledge.
5. Specific information about something.
6. Carnal knowledge.

We conclude therefore that within the scope of the Knowledge-Centered Web, it is necessary to deliver knowledge as enduring, justified and true consensus to users depending on context, users demands, desiderata and intention, whereby these aspects are supported by social facets, the environment, the profile, tasks and life cases of the users. Life cases, portfolios and tasks constitute the information demand of every users. The information demand of users requires a special content quality. It results in the requested knowledge, which is also depending on the understanding and motivation of users. So, the requested knowledge of users is a composition of understanding, and information demand, whereby the information demand is an aggregated component of life cases, motivation, intention and quality.

1.4. The Knowledge Delivery Task for Web 3.0 Knowledge Systems Environments

The knowledge delivery task of the Knowledge-Centered Web is defined as:
`Deliver the knowledge the user really needs through (1) concepts at the educational level level of the user that are illustrated and extended by (2) content which is quality content depending on the external and internal quality of the aggregated data (media object suite) and that are depicted by (3) topics in the language, in the culture and in the application portfolio of the user.`

Therefore, knowledge delivery and acquaintance for the user is user-oriented and life-case-based content, concepts and topics.

1.5. Survey of This Paper

Section 2 starts with a discussion of one of the most challenging applications for knowledge system environments: Cloud computing is based on highly distributed content, a large variety in understanding its concepts and in annotating content. It also provides functionality on demand. Therefore, the knowledge access is going to change to combination of knowledge chunks based on knowledge sharing. Beside the classical understanding of systems/infrastructure/software as a service we face also the research challenge of developing solutions for *knowledge-as-a-service*. Cloud knowledge sharing and provisions raises many novel research topics since the demand and the profiles of potential users and the capabilities of cloud services must be taken into consideration. One of the most challenging questions is the integration of context.

Section 3 approaches the development of knowledge system environment based on the observation that the carrier of knowledge is data. Therefore, knowledge services are based on *data-intensive services* that must provide knowledge based on the needs under direct supervision of data incompleteness and partnership context. A number of classical research solutions must be extended to such needs such as data or knowledge integration and collaboration support. Section

4 shows the potential and discusses the challenges of knowledge system environments to users. Knowledge processing is based on novel principles of *universal communication*. Since data is provided by different services with varying quality, varying scope and varying support, integration of data and compilation of knowledge must be enhanced by automatic quality maintenance. Creditability assessment is of the basic functions for knowledge compilation. The WISDOM system allows to derive creditability and reputation of web pages.

Section 5 develops a foundation for knowledge systems environments. This foundation is based on a separation of concern into content through data, concepts through models and notions, topics through language or carrier expressions and information through explicit representation of the user understanding. These dimensions can nicely be enhanced by a context dimensions that allows to develop cloud-based service systems as discussed in Section 2.

2. Cloud-Based Knowledge Services and Context Sensitivity

The recent development of IT has been towards service orientation, globalization and adaptability [26]. These demands have created opportunities for the development of cloud technology. According to [2] the development towards clouds is rooted in the development of Web 2.0 and the impact it has on the nature of the services offered. The shift from contractual and high-commitment services to low-commitment self-services has brought up the need for new kinds of knowledge sharing and storing opportunities.

2.1. Cloud Services

Cloud computing refers not only to applications delivered as services over the Internet, but also to the hardware and systems software that provide these services. The applications are usually referred to as SaaS, where the datacenter hardware and software are called the cloud [2] further define clouds as public, where a service is being sold in pay-as-you-go manner to the general public, and private clouds as the systems inside a company. They refer to public clouds as utility computing, thus defining cloud computing as the combination of SaaS and utility computing.

According to [2] the most relevant new aspects brought by cloud computing from a hardware point of view are: 1) the illusion of infinite computing resources available on demand, 2) the elimination of an up-front commitment by cloud users and 3) the ability to pay for use of computing resources on a short-term basis as needed. Utility computing is preferable when demand for service varies with time or when demand is unknown in advance.

[3] divide commercial cloud service architecture into four levels, which include physical machines, virtual machines, resource allocators and users. Services can be directed to each of these levels, which can be seen using the [40] ontology of five cloud service layers as described in Figure 2. The services are directed to applications, software environments, software infrastructure and hardware, end additionally to the software kernel, which consists of the physical servers that

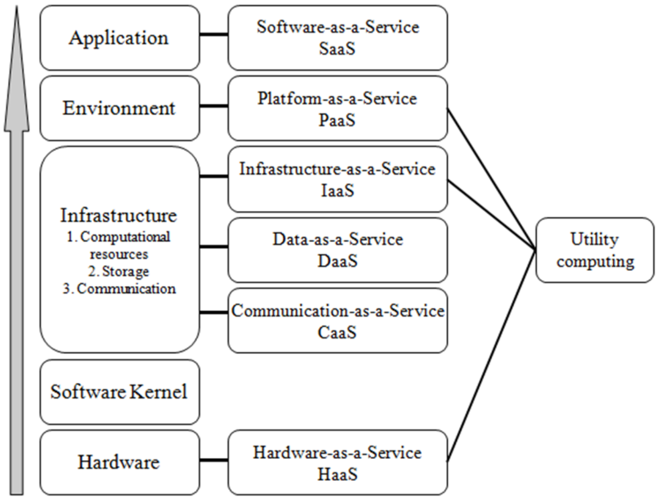

Figure 2. The cloud service layers and their services [40]

compose the cloud. The infrastructure layer consists of data, communication and computational resource services.

2.2. Using Clouds for Knowledge Sharing

Providing knowledge management tools as a service is a recent development in the IT service industry. The trend is towards a knowledge cloud to enhance the speed and accuracy of finding relevant knowledge for company use or customer service. Xu and Zhang (2005) state that, with a knowledge cloud type of service, the customer sends a query to a server and gets back an answer according to the data and knowledge mining specifications in the cloud.

For knowledge-sharing services, some new application opportunities are presented by [2]. For example, time consuming parallel batch processing and analytical tasks can be processed through an application that includes enough parallelism to hide the complexity of the task from the user, and simultaneously uses hundreds of cloud servers. The cost of moving large amounts of data into the cloud system must be evaluated in comparison to the need for speedy analysis. However, the need for data analysis is on the rise, as a growing share of computing resources is used for understanding customers, supply chains, buying habits, ranking and so on.

The knowledge-as-a-service involves three participants, as described in detail in the Figure 3. Data owners collect data from their daily business and are allowed to utilize the data in the cloud. They are themselves responsible for protecting the secrecy of the data outside the cloud. The knowledge service provider is divided into a knowledge extractor and an accessible knowledge server. Data sent by the data owners is extracted through security algorithms and agreed sets of grouping or filtering algorithms to enable data queries to hit the right data for each customer.

[41] accentuate the fact that the knowledge server owns the knowledge, but does not necessarily own the data behind it. Service providers serve the customers who consult the knowledge base in their decision making. The customers can hide the query instance from the service provider to further enhance the security of their knowledge.

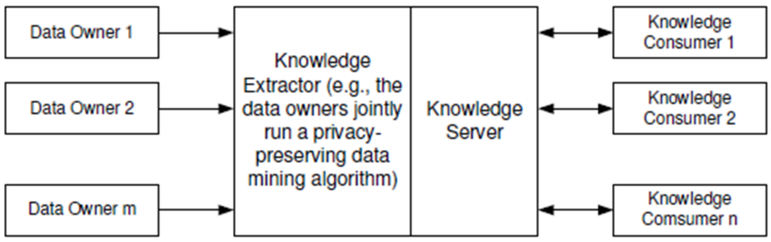

Figure 3. The paradigm of knowledge-as-a-service[41]

2.3. Problems of Using Clouds for Knowledge Sharing

Companies employ cloud services today to improve scalability and to react to varying resource demand. At the moment the services are often limited to one provider at a time due to inflexibility and varying interfaces [3]. In order to widen the use of clouds in services, some common standard interfaces have to be followed [3,40]. The lack of common solutions and regulations prevent the companies from utilizing the open market data processing providers; according to [3] this also poses a threat to the ability to develop market-oriented knowledge services.

Cloud technology is improving quickly, but currently there are no common guidelines or policies. This creates issues of a legal and ethical nature regarding knowledge sharing. The security of the knowledge is of the upmost priority of many organizations, but there is no common legislation to cover the issue, even less if the cloud is international. Knowledge breaching (attack on the server) is one of the knowledge security issues connected to centralizing data in large service provider entities. Using a small number of adaptively selected queries, the attacker can use the responses to pry into parts of the knowledge. Combining these parts can generate knowledge close to the original.

2.4. Context Sensitivity of Knowledge and Clouds

Cloud computing only works if the data flow from the user to the cloud, within the cloud and back to the user is efficient. Geographic, technical and political borders can fracture the cloud into smaller pieces and hinder the distribution of knowledge. According to [14] it is also important to consider how the user can extract certain information according to the context. Context is any information that can be used to characterize the situation of a person, place or object relevant to the interaction between a user and an application. This context definition by [4] also includes the user and application themselves.

A communication situation usually has a context, and as [14] state, context is expressed as associations between knowledge units. Without context, new knowledge will not necessarily be absorbed, understood, and accepted.

In the case of the knowledge-as-a-service model by [41], the knowledge service provider acts as an intermediary between the input knowledge and the query responses of the customer, thus creating a context-sensitive environment for knowledge transfer. The response depends partly on the input data, but also on the filtering and grouping algorithm of the data extractor and the way the queries are made by the customers. When receiving the knowledge service provider's response, the customer will be further dependent on the context to interpret and absorb the knowledge, and to act accordingly.

The context sensitivity of the cloud as a means to distribute knowledge is manifested in all of the knowledge-as-a-service model stages, as there might be multiple actors handling the data inputs from the owners and knowledge outputs to the customers. As stated by [17, p. 238], the absorption of modern techniques has virtually nothing to do with deep-rooted core beliefs. We standardize for convenience, but the mental agendas remain inviolate and hidden.

2.5. Future

The infrastructure software of the future will probably run on virtual machines. Application software will most probably be divided into the customer part and the cloud part. Of these, current Web 2.0 systems do not yet offer utilization of the customer application part when it is not connected to the cloud. The cloud part itself has to be very efficiently scalable both up and down, which is not typical in the case of traditional software systems. On the hardware system level, new clouds require development of memory hierarchy, as the lowest level of software will be virtual machines. Also, routers, switches, and bandwidth will need improvements to meet the needs of the future [2].

Globalization is one of the main trends in the IT world. In the increasingly geographically distributed IT environment, clouds are becoming tools that allow even more possibilities to collect and share knowledge between actors all over the world. Globalization and the use of cloud services increase the impact of context in knowledge sharing, bringing into consideration the effect of national cultures. Current cloud interfaces, as mentioned above, have to be standardized in order to enable open cloud service markets. If the context sensitivity and knowledge security issues can be acknowledged and tackled, cloud computing could be the future of knowledge distribution and services.

3. Towards Knowledge Systems for Data-Intensive Science

Knowledge is seen as human understanding gained through study and experience, which includes perception, skills, training, common sense, and experience. Traditional knowledge systems are passive archives of information on facts and heuristics, but pay little concern for the context in which they will be used. Meanwhile, knowledge systems are expected to provide actionable information available in the right format, at the right time and at the right place for decision making. Therefore, next-generation knowledge systems need to evolve to be a more sophisticated platform for evaluating validating and presenting information content with respect to verifiable truth claims.

A motivating example is shown in the evolution from experimental and theoretical sciences to emerging data-intensive science [12]. Today, scientists are exploring data captured by large-scale, complex instruments or generated by simulations stored in computers. In this emerging phase, knowledge developed primarily for the purpose of scientific understanding is being complemented by knowledge created to target practical decisions and action. Until now, the demand for applications science is emerging. Followings show the core characteristics of the applications science compared with the basic science:

Need driven. The applications science focuses on the knowledge to seek courses of action and determine their consequences (i.e., societal needs), rather than seeking answers to questions (i.e., scientific curiosity).

Externally constrained. External circumstances dictate when and how applications knowledge is needed, not according to academic schedules based on when and how the best knowledge can be obtained.

Useful even when incomplete. Despite incomplete data or partial knowledge due to the loss of scientific stationarity, the means of making effective use of partial knowledge must be developed, including robust inferences and statistical interpretation, in order to make decisions with establishing confidence.

Addressing the above challenges, interdisciplinary synthesis of knowledge resources is a crucial requirement for next-generation knowledge systems. Figure 4 shows a practice in National Institute of Information and Communications Technology (NICT) for space environment science [20]. The system aims to find associations among heterogeneous domains of data, such as solar-space observation data, space weather forecast data, biology data, science chronological data, earth weather station data, and newspaper article data. The principal questions of users reflect societal needs rather than scientific curiosity, for instance "What are the social implications of ionospheric plasma bubbles?", "To what extent are sunspot activities related to economic cycles?". To answer those questions, the system analyzes correlations (both semantic and spatiotemporal) among the multidisciplinary data, and makes inferences to find a set of associative data as an evidence for the questions, or user's hypotheses.

This kind of knowledge system introduces following unique challenges :

Data intensive. Data sources that support basic science are often insufficient to support applications. New applications-appropriate sources must be identified, and new ways of observing (including the use of communities as data gatherers) must be developed. While science data has value for data-owner scientists, sharing data with others should increase rather than decrease that value.

Service-oriented architectures. Service orientation is widely acknowledged for its potential to revolutionize the world of computing by abstracting from all resources as services in a service-oriented architecture (SOA) [24]. The SOA

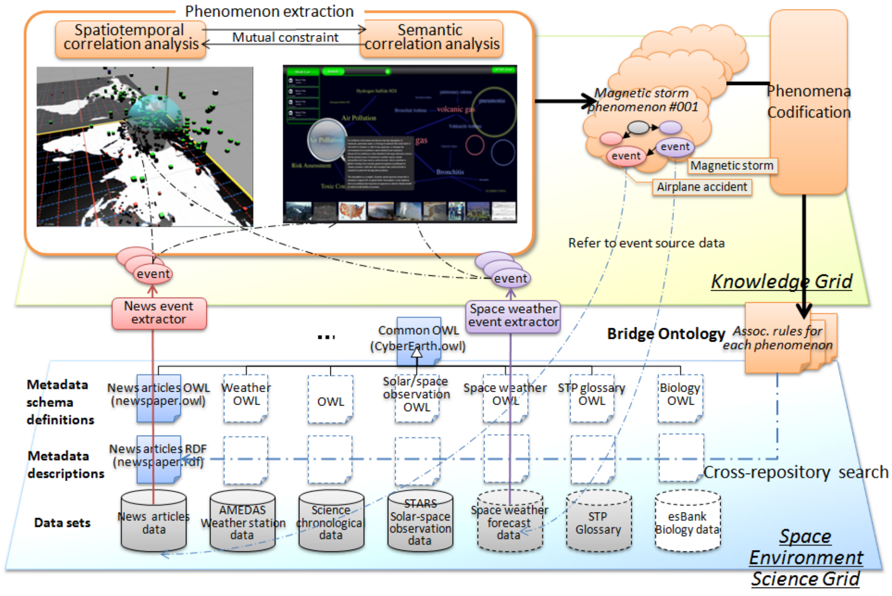

Figure 4. The knowledge system discovering phenomenon-oriented relationships among heterogeneous domains of data in space environment science.

helps to mitigate the transition to new underlying technologies and enable the linkage of data and resources.

Migration to cloud computing. Cloud computing allows to host, process, and analyze large volumes of multidisciplinary data. It offers obvious advantages, such as co-locating data with computations and an economy of scale in hosting the services. Studies on large synthesis datasets also create a need for collaborative tools in the cloud.

3.1. Knowledge Correlation Computing

Large synthesis of multidisciplinary data builds an evolving network of community knowledge. NICT's knowledge cluster systems [42] aim to realize it by linking heterogeneous knowledge resources owned by different communities. The basic idea is similar as the World Wide Web (the Web in short). As the Web provides a framework of infinitely-evolving document repository, the knowledge cluster systems provide a framework of infinitely-evolving knowledge repository. Because knowledge is context-dependent information in nature (remember that knowledge is actionable information available in the right format, at the right time and at the right place for decision making), links between knowledge resources are also context-dependent in contrast with the simple links between Web pages in conventional Web.

Linking two different knowledge resources requires "bridge concept", the information describing semantic relations between two different communities of knowledge. We are proposing a method for organizing the bridge concepts by using the semantic space based on vector space model, in which the knowledge is rep-

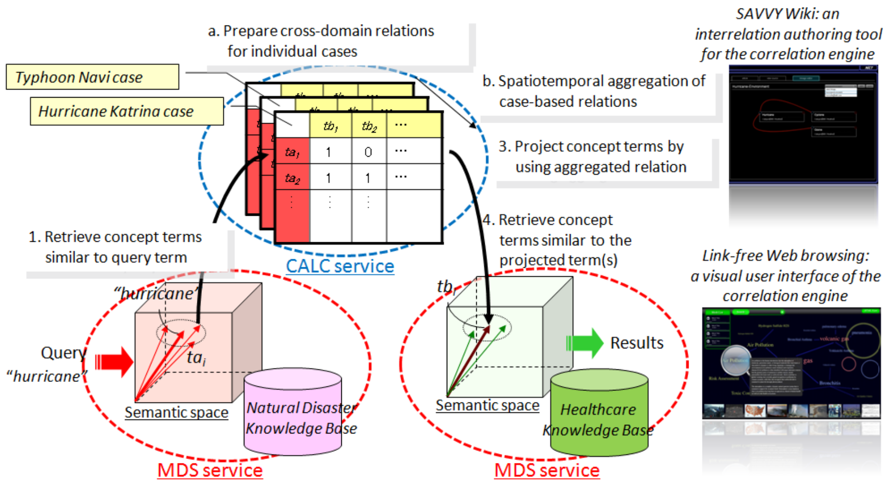

Figure 5. Knowledge correlation computing by semantic space model.

resented by a vector in the subspace (Figure 5). The advantage of this method is that we can manage context-dependent correlations between knowledge resources by measuring vector correlations in subspaces [21]. In addition, time scale or geographical scale can be included in the context. The link mechanism focuses on the relations between knowledge resources only in the given context, by which the complexity of bridge concept is greatly reduced compared with giving universal relations in traditional approaches. The relations are defined as significance of correlations in liner scale. It allows ambiguity or uncertainty of relations between different domains of knowledge, which increase flexibility of the bridge concept.

3.2. Data-centric Service Collaboration on Global Knowledge Grid

The Global Knowledge Grid [43]is an ICT infrastructure for implementing the knowledge cluster systems based on collaborations of knowledge services [23] in distributed environments. The Global Knowledge Grid stems from the necessity of providing knowledge and processing capabilities to everybody, thus promoting the advent of a competitive knowledge-based economy. Knowledge services on the Global Knowledge Grid are classified into the following categories: **(1) knowledge discovery service** provides concept descriptions of an information source based on knowledge discovery approach by data mining, such as segmentation, classification, summarisation, and ontological annotations. The MDS services in Figure 5 are examples of this type of service. (2) **knowledge association service** produces various kinds of associations between multiple information sources based on their concept descriptions provided by the corresponding knowledge discovery services. The CALC service in Figure 5 is an example of this type of service. **(3) knowledge delivery service** is responsible for (visual) presentation and navigation of the results from knowledge discovery services and knowledge association services. It also handles user interfaces and interactions. These types of knowledge services are developed and deployed by a wide variety of groups joining the Global Knowledge Grid in parallel.

An application on the Global Knowledge Grid is realized by a specific collaboration of those services. The Global Knowledge Grid provides a data-centric service collaboration, called Service MeshUp. Figure 6 illustrates how the Global Knowledge Grid works according to the Service MeshUp mechanism. The Service MeshUp, described by our original XML-based process definition language, consists of a set of aspects, each of which defines a single task or function in an application. Each aspect has its own properties (aspect properties) to be shared by the services. For each aspect, an application developer determines a set of services to be involved, and then, for each service, he/she defines precondition, behavior, and postcondition with respect to the aspect properties. Here, the precondition describes the conditions for activating the service with respect to the aspect properties. For instance, the precondition becomes true when a specific aspect property is modified, or has a specific value. The behavior defines what the corresponding service will do when it is activated. Basically, the functions of the corresponding service are invoked. The postcondition defines what will be done after the behavior. In most cases, the data obtained as the result of the behavior will be set to the aspect properties. In contrast with conventional service mashup methods based on workflow model, like WS-BPEL [8], our Service MeshUp aims at realizing service collaborations sharing common interests, each of which is represented by aspect properties. While both models are convertible with each other, our Service MeshUp model is more suitable to self-organization processes and ever-evolving processes such as building collective knowledge or monitoring situations than transactional processes like business workflows.

4. Knowledge Processing for Universal Communication

4.1. Universal Communication

Nowadays, the Internet has become an essential information infrastructure. We search information on the Internet when we need to obtain knowledge. Furthermore, ubiquitous and mobile computing technologies are providing a new knowledge processing technology in the real-world.

The ICT infrastructure which is capable of distributing massive amounts of information anytime and anywhere, is becoming more wide spread. The infrastructure supports human communication in order to provide useful information. Universal communication, any kind of information with anyone, is a big challenge of a next generation information analysis technology. Some of barriers are still remained to realize universal communication: language barrier, miscommunication due to untrusted information and long distance. Those barriers should be overcome by effective new technology. We are focusing on information analysis technology to find valuable and credible information from Web content.

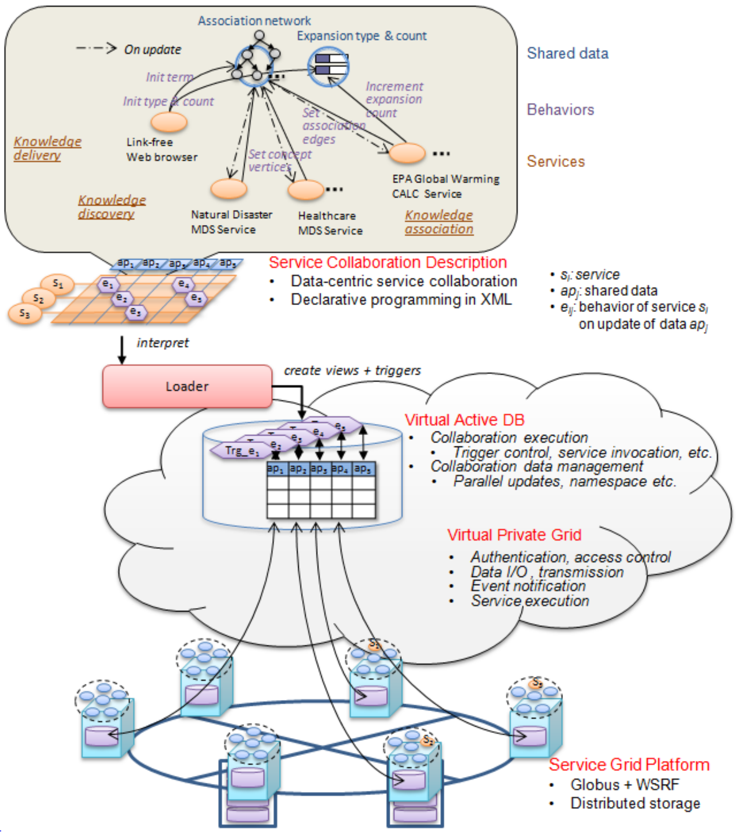

Figure 6. Global Knowledge Grid: A data-centric SOA platform for synthesis of knowledge resources.

4.2. Information Analysis for Credibility of Web Content

General users write their own daily news and post information they consider interesting as digital documents for blogs and SNS. Such digital content includes both valuable information and worthless, false, and demagogic information. Ordinary web search engines can display web pages in a particular order. The top-ranked web page does not always contain valuable and credible information. Nevertheless, readers often trust the authenticity of the displayed information. Web search engines such as Google retrieve web content using search keywords. The PageRank method evaluates the score of web content and generates a ranked list. The top-ranked web content on search engines is often relevant to the user's query, though, in some cases, the content may not be credible or valuable. Even if users believe that the content is useful, the search engine cannot evaluate the retrieved digital content, and users have to retrieve various contents using different keywords.

NICT's Information Analysis Technology addresses the issue of information credibility by analyzing credibility based on the following criteria: 1) content, 2) sender, 3) appearance, and 4) authenticity of the content. We believe that the understanding of texts by a machine is important, and that an NLP approach is very effective in evaluating the credibility criteria. By using different methods for analyzing the information credibility criteria, credible information can be obtained, which eventually becomes valuable knowledge. Sometimes, vast amounts of knowledge may have to be combined to understand even a single social topic, which is not feasible when using conventional search engines.

An Ordinary Web search engine does not consider the credibility of the Web content. The top-ranked Web content corresponds to the level of credibility the search engine displays. A novel technology that extracts information credibility criteria from the digital content on the Internet is expected to be developed as a next-generation Web technology.

We strongly believe that the understanding of text by a machine is vital for the information credibility criteria. In the conventional document-processing method, a document is considered as a mere bag of words. NICT's approaches are based on the natural language processing (NLP) technology. NICT's ultimate goal is to develop information analysis methods to evaluate the four above mentioned criteria. The method will show multifaceted credibility criteria for supporting the human judgment of information credibility. The aim of this study is to make the selection and distribution of credible information easier by using our analysis methods. People use the Web not only to obtain information on what they want to know but also to collect information for decision making. They generally use conventional search engines such as Yahoo! and Google, but it is difficult to judge the credibility of information from the obtained search results. This is because these search engines rank Web pages in terms of a single measure that is not always related to the credibility of the Web pages. To judge the credibility, users must examine the Web pages individually. For example, when a user searches for information on "bio-ethanol", which is arguably said to be a bio-based fuel, using a conventional search engine, commercial pages that advertise its effect on environment are highly ranked, and the other pages remain in low ranks. From this search result, it is difficult to judge whether "bio-ethanol" is really good for environment. In order to obtain credible criteria information, it is necessary to develop a new system that automatically processes information on a given topic and provides a bird's eye view and multifaceted views on the topic. For example, these views are a summary page that tabulates reports and opinions on the topic, people and organizations that wrote the pages, whether contact addresses and information sources are specified, and so forth.

We have been developing an information credibility analysis system, WISDOM (Web Information Sensibly and Discreetly Ordered and Marshaled) [1]. WISDOM enables users to browse a large volume of information from a bird's eye view and multifaceted views while changing a search condition on the basis of each factor of information credibility. By using this system, users can obtain credible information on a topic of interest more efficiently.

We propose a method for providing a bird's eye view of major statements on a given topic and their contradictions. This method targets Web pages on the topic, and extracts linguistic expressions occurring with a high frequency as major statements; this method also extracts contradictions to the major statements.

This resulting view is a summary that enables users to grasp what facts and opinions are found on the Web. In this summary, not only major statements but also their contradictions are presented even if the number of the contradictions is small. That is to say, this system can extract statements that are related to a given topic including minor ones, and provides users with a bird's eye view of the topic.

4.3. The overview of WISDOM

The purpose of WISDOM is to evaluate the credibility of information on the Web from multiple viewpoints. WISDOM considers the following to be source of information credibility: information contents, information senders, and information appearances. We aim at analyzing and organizing these measures on the basis of semantics-oriented natural language processing (NLP) techniques.

The analysis of information contents focuses on the text information in Web pages and also includes the processes of page clustering,summarization, and opinion extraction. The analysis of information senders highlights the identity of a sender, classifies his/her affiliation, and determines the level of expertise in the topic. The analysis of information appearances focuses on the impression of a Web page, and includes the processes of clarifying the information source and contact address and assessing the appropriateness of page design and writing style.

WISDOM is designed to function on several hundred million Web pages, and currently targets 120 million Japanese Web pages. For each Web page,sentence extraction, morphological and syntactic analysis, opinion extraction [22], classification of the information sender, and assessment of information appearance are performed in advance. These results are stored in an XML-based format.

A user accesses WISDOM via a Web browser, and inputs a topic that he/she wants to know as a query. By clicking analysis button, he/she can see the results organized and summarized according to the information senders, contents,and opinions. These results represent an overall trend of the topic (“information sender analysis”, “page clustering” and “opinion distribution” . The user can also see the detailed information of each Web page, such as the information sender (author) of the page, the appropriateness of page appearance, and individual opinions with their positions among the entire opinion distribution. In this way, the user can browse information of the topic from various points of view and find credible information more efficiently.

4.4. Universal Communication using Web Content

NICT's Information Analysis Technology can analyze social reputation on Web pages. The distribution of reputation enables us to find more useful information when we visit unfamiliar places. For examples, mobile phone can search some restaurants around there when we want to go good restaurant in the place. NICT is developing Spoken Dialog System. The system is based on mobile computer and mobile phone. Users can input query by voice. The system has a speech recognition/synthesis and a dialog management function. Speech recognition function transfers from voice query to text query. Dialog function manages user's situa-

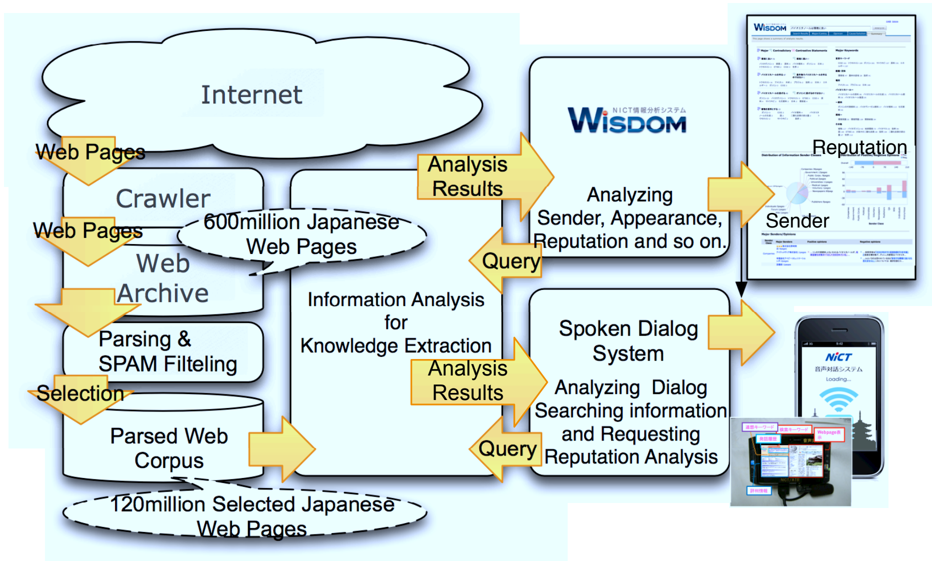

Figure 7. Overview of NICT's Information Analysis Infractructure.

tion and context. The system communicate with information analysis function in order to retrieve related information and analyze reputation on Web pages (See Fig 7) . Spoken dialog system supports user's activities anywhere and anytime. The answer is output by synthesis speech. Users can overcome operation barrier of information retrieval. It is one of ICT supported system for the Universal Communication.

5. Foundation of Knowledge Systems by Knowledge Chunks

Knowledge can be characterised through (1) its content, (2) its concepts, (3) its annotations or topics, and (4) its understanding by the user. Knowledge pieces cannot be considered in an isolated form. For this reason we imagine to use knowledge chunks as a suite of knowledge pieces consisting of content, concepts, topics and information. These dimensions are interdependent from each other. Figure 8 displays the knowledge space.

5.1. Content and Media Types: The Data Dimension.

Content is complex and ready-to-use data. Content is typically provided with functions for its use. Content can be defined n the basis of *media types*. Content management systems are information systems that support extraction, storage and delivery of complex data.

Content in its actual definition is any kind of information that is shared within a community or organization. In difference to data in classical database systems

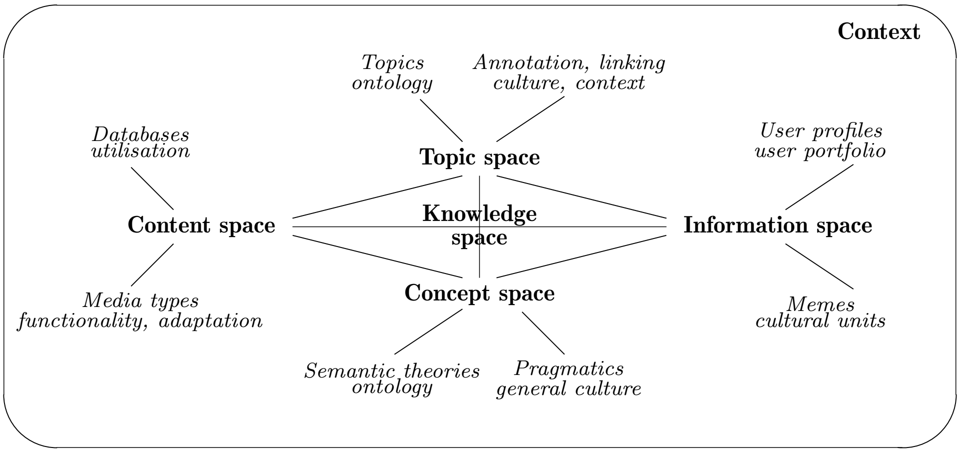

Figure 8. The four dimensions of the knowledge space surrounded by the context dimension: (1) data dimension through content; (2) foundation dimension through concepts; (3) language dimension through topics; (4) user dimension through information; (5) context of data (content) or theories (concept) or user (information) or carrier/language (topic)

content usually refers to aggregated macro data which is complex structured. Structuring of content can be distinguished:

- The structure of the aggregated micro data is preserved but micro data was combined to build larger chunks of information. Examples are scientific data sets such as time series of certain measurements. There is a common (or even individual) structuring and meaning for each sampling vector but the compound of all sampling vectors adds additional semantics.
- The structure of content is only partially known. A typical example is the content of Web pages: structuring is known up to a certain level of detail which may also be varying within one instance.
- Content may be subsymbolic, such as pictures, videos, music or other multimedia content.

Aggregation of content usually takes place by combining reusable fragments provided by different sources in different formats such as texts, pictures, video streams or structured data from databases. Content is subject to a content life cycle which implies a persistent change process to the content available in a content management system (CMS).

The more generic ones agree in a major paradigm: the separation of data management and presentation management. Data management reflects the process of supporting content creation, content structuring, content versioning, and content distribution while presentation management grabs the data for delivering it to the user in various ways. Only content which is generated following this separation can be easily shared, distributed, and reused.

Following new trends and developments in Web technologies, e.g., in the context of Web 2.0 or the Semantic Web the automated processing of content becomes more and more important. Because content represents valuable assets it may be reused in different contexts (*content syndication*) or has to remain accessible for a long time.

The semistructured or even unstructured nature of content requires annotations to enable search facilities for content. Expressing semantics in a machine interpretable way has been under investigation since the early days of artificial intelligence, see e.g., [36] for a survey of knowledge representation techniques such as logical theories, rule-based systems, frames or semantic nets. Today systems handle semantical descriptions as metadata describing certain content instances. There are different ways for associating data and metadata:

- A conceptual, logical, or physical *schema* is defined and instances are created according to this schema. This is the usual way for classical databases. The modelling language strongly restricts the capabilities of this description facility. Common languages such as Entity-Relationship Modelling or UML focus on structural properties with support of selected integrity constraints.
- Defining a schema is not applicable (or only in a restricted way) to semistructured or unstructured content. For that reason content instances are annotated. An annotation is a triple (S, P, O) where S denotes the subject to be annotated, P a predicate denoting the role or purpose of this annotation, and O the object (or resource) which is associated with S. The vocabulary for annotations is organized in ontologies and thesauri. A typical language for expressing annotations in the context of the Semantic Web is the Resource Description Framework (RDF, [39]) while the Web Ontology Language OWL ([38]) may be used to express semantic relationships between the concepts and resources used for annotation. There exist myriads of ontologies and parameter definitions for different application domains such as the Dublin Core parameters [5]) for editorial content.

5.2. Concepts and Theories: The Foundation Dimension.

Concepts are the basis for knowledge representation. They specify our knowledge what things are there and what properties things have. Concepts are used in everyday life as a communication vehicle and as a reasoning chunk. Concepts can be based on definitions of different kinds. Therefore our goal for the development of knowledge web can only be achieved if the content definition covers any kind of content description and goes beyond the simple textual or narrative form.

A general description of concepts is considered to be one of the most difficult tasks. We analysed the definition pattern used for concept introduction in mathematics, chemistry, computer science, and economics. This analysis resulted in a number of discoveries:

- Any concept can be defined in a variety of ways. Sometimes some definitions are preferred over others, are time-dependent, have a level of rigidity, are usage-dependent, have levels of validity, and can only be used within certain restrictions.
- The typical definition frame we observed is based on definition items. These items can also be classified by the kind of definition. The main part of the definition is a tree-structured structural expression of the following form

SpecOrderedTree(StructuralTreeExpression
(DefinitionItem, Modality(Sufficiency, Necessity),
Fuzziness, Importance, Rigidity,
Relevance, GraduationWithinExpression, Category))) .

- Concepts typically also depend on the application context, i.e. the application area and the application schema. The association itself must be characterised by the kind of association.

Concepts are typically hierarchically ordered and can thus be layered. We assume that this ordering is strictly hierarchical and the concept space can be depicted by a set of concept trees. A concept is also dependent on the community that prefers this concept. A concept is also typically given through an embedding into the knowledge space. The schema in Figure 9 displays the general structure for content definition. This schema also covers all aspects discussed in [19]. A concept has

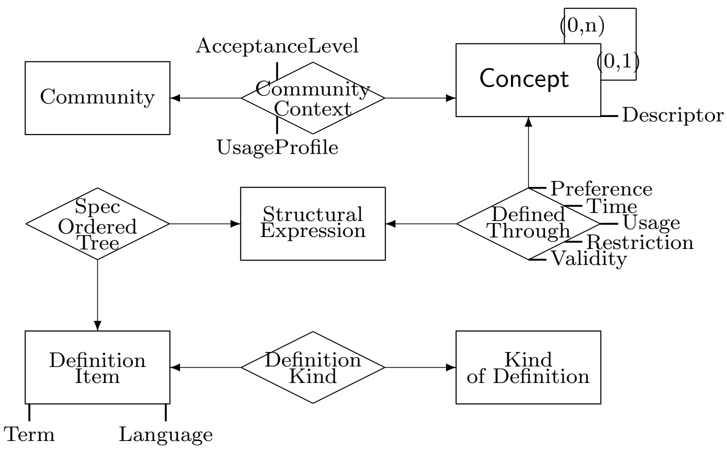

Figure 9. The main schema for Concept Definition and Formation

typically a manyfold of definitions. Their utilisation, exploration and application depend on the user (e.g. the education profile), the usage, and context.
Example (concept of mathematical set): *A set can be given by an enumeration of its elements, by inductive definition of its elements, by an algorithm for the construction of the set, or by explicit description of the properties of the set. Which of the definitions is more appropriate depends on the application domain.*

These set definitions are based on the principle of extensionalism: Two sets are equal if they contain the same elements. We might also use sets that are not based on this principle.

5.3. Topics and Ontologies: The Language Dimension.

Content and concepts may be enhanced by **topics** that specify the pragmatic understanding of users.

Semantic annotation in current content management systems is usually restricted to preselected ontologies and parameter sets. Rich conceptual data models are only available in more sophisticated systems. Because most generic CMS are focused on Web content management semantic annotation is usually restricted to editorial parameters. Specialized content management systems which are adapted to certain application domains incorporate preselected and tailored ontologies. Especially for XML-based content there exist several annotation platforms which incorporate semantical annotation either manually or semi-automatically; see [25] for a survey on available platforms.

Automated processing of semantical metadata is usually restricted to search facilities, e.g., searching for the author of an article. Because ontologies are preselected for most systems a full-featured reasoning support is usually not available. Especially for OWL ontologies there are reasoning tools based on description logics such as Racer ([10]) or FaCT which enable T-box (but also A-box) reasoning about semantic relationships between annotation concepts.

Applying generic semantical annotation and classical reasoning facilities to content management suffers from several drawbacks:

- Content as aggregated macro data is only partially analysable. The purpose of metadata is the description of properties which cannot be concluded from the data itself. The very simple annotation frame of (S, P, O) triples does not allow one to express complex properties. For that reason this information has to be kept in the underlying ontology by defining appropriate concepts. The support of user-specific concepts increases the size of the ontology significantly and makes reasoning support even harder. Ad hoc definitions of user-specific concepts is not supported in this annotation model.
- Annotation with respect to arbitrary ontologies implies general purpose reasoning support by the system. Reasoning for even simple languages suffers from its high computational complexity (e.g., NEXPTIME for the restricted OWL-DL dialect, [13].) Dealing with high worst-case complexities implies a small size of input data but this is a contradiction to expressible ontologies and the definition of content as complex structured macro data. Especially the size of content instances is a crucial factor because A-box reasoning is a critical point for automated content processing ([11].)

But there are advantages, too:

- Usually, it is possible to distinguish between different points of view on content instances. Not every property is important while looking from every point of view. The macro data may encapsulate and hide properties from its aggregated micro data. Reasoning about the properties of the compound can be separated from the properties of the elements as well as the properties of interconnections between content instances.
- Typical application scenarios determine important properties and suggest evaluation strategies. So ontologies may be decomposed to enable a contextualized reasoning, e.g., on the basis of Local Model Semantics ([9]). Local reasoning may rely on a language that is just as expressive as needed in this context. Contexts relying on less expressive languages may support automated reasoning while contexts relying on more expressive languages may be used for manually interpreted information. Soundness and completeness

of the reasoning process are not of primary interest as long as the reasoning result is acceptable in the application domain.
- The separation between annotations relying on common knowledge, user-specific annotations and (especially) usage-specific annotations reduces the size of incorporated ontologies significantly.
- If semantic annotations themselves are given a more sophisticated internal structure reasoning can be adapted to the requirements of the application domain.

The major disadvantage of current semantic description in content management is the treatment of knowledge over content instances as metadata on a secondary level in a strongly restricted language. In the following sections we will introduce a data model for content which handles the semantic part on the same level as the content itself and gives additional structure to the semantic description. Content chunks are semantically enriched content instances. They are based on the notion of a schema for content chunks to incorporate typical functionality of content management systems such as content generation, content delivery, or content exchange.

5.4. Information and Memes: The User Dimension.

There are several definitions for information[7]. We categorize these notions:

- The first category of these definitions is based on the mathematical notion of entropy. This notion is independent of the user and thus inappropriate in our project context.
- The second category of information definitions bases information on the data a user has currently in his data space and on the computational and reasoning abilities of the user. Information is any data that cannot be derived by the user. This definition is handy but has a very bad drawback. Reasoning and computation cannot be properly characterised. Therefore, the definition becomes fuzzy.
- The third category is based on the notion of information utility. Business information systems understand information as data that have been shaped into a form that is meaningful, significant and useful for human beings.

[7] In general, information is
- raw data and
- well-formed and meaningful data
- that (1) has been verified to be accurate and timely relative to its context,
  (2) is specific and organized for a purpose,
  (3) is presented within a context that gives it meaning and relevance, and which
  (4) leads to increase in understanding and decrease in uncertainty.

This notion extends the GDI notion (General Definition of Information). "Well-formed" means that the raw data are clustered together correctly, according to the rules (syntax) that govern the chosen system, code or language being analysed. Syntax is understood broadly, as what determines the form, construction, composition or structuring of something. "Meaningful" means that the data must comply with the meanings (semantics) of the chosen system, code or language in question. We refer to [35] for different kinds of semantics. However, let us not forget that semantic information is not necessarily linguistic. For example, in the case of the manual of the car, the illustrations are such as to be visually meaningful to the reader.

These data satisfy an information demand and can be understood by this group. Typical data represent information about significant people, places, and things within an organisation or in the environment surrounding it.
- The fourth category is based on the general language understanding of information [27]. Information is either the communication or reception of knowledge or intelligence. Information can also defined as

  ∗ knowledge obtained from investigation, study, or instruction, or
  ∗ intelligence, news or
  ∗ facts and data.

  Information can also be the act of informing against a person.
  Finally information is a formal accusation of a crime made by a prosecuting officer as distinguished from an indictment presented by a grand jury.

All these definitions are too broad.

We are thus interested in a definition that is more appropriate for the internet age (*anthroposophic* understanding) in extension of the GDI notion of information.

Information as processed by humans,

- is carried by *data*
- that is perceived or noticed, selected and organized by its receiver,
- because of his subjective human interests, originating from his instincts, feelings, experience, intuition, common sense, values, beliefs, personal knowledge, or wisdom,
- simultaneously processed by his cognitive and mental processes, and
- seamlessly integrated in his recallable knowledge.

The value of information lies solely in its ability to affect a behavior, decision, or outcome. A piece of information is considered valueless if, after receiving it, things remain unchanged. For the technical meaning of information we consider the notion used in information theory.

Therefore, information is directed towards pragmatics, whereas content may be considered to highlight the syntactical dimension. If content is enhanced by concepts and topics, then users are able to capture the meaning and the utilisation of the data they receive. In order to ease perception we use *metaphors*. Metaphors may be separated into those that support perception of information and into those that support usage or functionality.

Users are reflected by actors that are abstractions of groups of users. Pragmatics and syntactics share data and functions. The functionality is provided through functions and their representations. The web utilisation space depends on the technical environment of the user. It is specified through the layout and the playout. Layout places content on the basis of a data representation and in dependence of the technical environment. Playout is based on functionality and function representations, and depends on the technical environment.

The *information transfer* from a user A to a user B depends on the users A and B, their abilities to send and to receive the data, to observe the data, and to interpret the data. Let us formalise this process. Let s_X denote the function

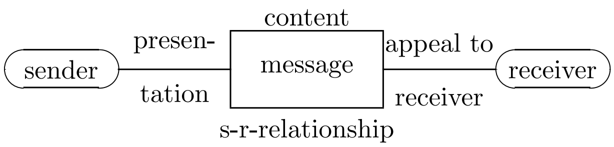

Figure 10. Dimensions of understanding messages

user by a user X for data extraction, transformation, and sending of data. Let r_X denote the corresponding function for data receival and transformation, and let o_X denote the filtering or observation function. The data currently considered by X is denoted by D_X. Finally, data filtered or observed must be interpreted by the user X and integrated into the knowledge K_X a user X has. Let us denote by i_X the binary function from data and knowledge to knowledge. By default, we extend the function i_X by the time t_{i_X} of the execution of the function.

Thus, the data transfer and information reception (or briefly information transfer) is formally expressed it by

$$I_B \;=\; i_B(o_B(r_B(s_A(D_A))), K_B, t_{i_X}) \;.$$

In addition, time of sending, receiving, observing, and interpreting can be taken into consideration. In this case we extend the above functions by a time argument. The function s_X is executed at moment t_{s_X}, r_X at t_{r_X}, and o_X at t_{o_X}. We assume $t_{s_A} \leq t_{r_B} \leq t_{o_B} \leq t_{i_B}$ for the time of sending data from A to B. The time of a computation f or data consideration D is denoted by t_f or t_D, respectively. In this extended case the information transfer is formally expressed it by

$$I_B \;=\; i_B(o_B(r_B(s_A(D_A, t_{s_A}), t_{r_B}), t_{o_B}), K_B, t_{i_B}) \;.$$

The notion of information extends the dimensions of understanding of message displayed in Figure 10 to a web communication act that considers senders, receivers, their knowledge and experience. Figure 11 displays the multi-layering of communication, the influence of explicit knowledge and experience on the interpretation.

The communication act is specified by

- the communication message with the content or content chunk, the characterisation of the relationship between sender and receiver, the data that are transferred and may lead to information or misinformation, and the presentation,
- the sender, the explicit knowledge the sender may use, and the experience the sender has, and
- the receiver, the explicit knowledge the receiver may use, and the experience the receiver has.

We approach the analysis of knowledge system usage as the first important part of storyboarding pragmatics. Knowledge system usage analysis consists of three parts:

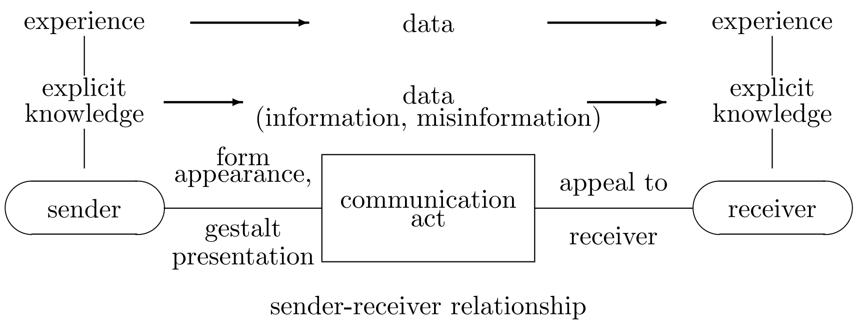

Figure 11. Dimensions of the communication act

1. *Life cases* capture observations of user behaviour in reality. They can be used in a pragmatic way to specify the story space. The work on life cases was reported in a previous publication [33].
2. *User models* complement life cases by specifying user and actor profiles, and actor portfolios. The actor portfolios are used to get a better understanding of the tasks associated with the knowledge system. The work on user models was reported in a previous publication [34].
3. *Contexts* complement life cases and user models by characterising the situation in which a user finds him/herself at a certain time in a particular location. We classify various aspects of contexts related to actors, storyboard, system and time, which make up the context space, then analyse each of these aspects in detail. This is formally support by lifting relations.

User modelling is based on the specification of *user profiles* that address the characterisation of the users, and the specification of *user portfolios* that describe the users' tasks and their involvement and collaboration on the basis of the mission of the knowledge system [30].

To characterize the users of a knowledge system we distinguish between *education*, *work* and *personality* profiles. The education profile contains properties users can obtain by education or training. Capabilities and application knowledge as a result of educational activities are also suitable for this profile. Properties will assigned to the work profile, if they can be associated with task solving knowledge and skills in the application area, i.e. task expertise and experience as well as system experience. Another part of a work profile is the interaction profile of a user, which is determined by his frequency, intensity and style of utilization of the knowledge system. The personality profile characterises the general properties and preferences of a user. General properties are the status in the enterprise, community, etc., and the psychological and sensory properties like hearing, motoric control, information processing and anxiety.

A *portfolio* is determined by responsibilities and is based on a number of targets. Therefore, the actor portfolio (referring to *actors* as groups of users with similar behaviour) within an application is based on a set of tasks assigned to or intended by an actor and for which s/he has the authority and control, and a description of involvement within the task solution [32]. A *task* as a piece of work is characterized by a problem statement, initial and target states, collaboration and presupposed profiles, auxiliary conditions and means for task completion. Tasks may consists of subtasks. Moreover, the task execution model defines what, when, how, by whom and with which data a task can be accomplished. The result of executing a task should present the final state as well as the satisfaction of target conditions.

For task completion users need the right kind of data, at the right time, in the right granularity and format, unabridged and within the frame agreed upon in advance. Moreover, users are bound by their ability to verbalise and digest data, and their habits, practices, and cultural environment. To avoid intellectual overburdening of users we observe real applications before the system development leading to *life cases* [33]. Life cases help closing the pragmatic gap between intentions and storyboarding. They are used to specify the concrete life situation of the user and characterise thus a bundle of tasks the user should solve. Syntax and semantics of life cases have already been well explored in [30].

In addition, each user has an *information portfolio*, which specifies the information needs as well as the information entered into the system. We do not model the information portfolio as part of a user, but instead of this we will model the information "consumed" and "produced" with each more detailed specification of a user request.

5.5. Context Dimension Characterisation and Adaptation of Knowledge Delivery by Context

Taking the commonly accepted meaning a context [15] characterises the situation in which a user finds him/herself at a certain time in a particular location. In this sense context is usually defined only statically referring to the content of a database. Only very few attempts have been made so far to consider context of scenarios or stories.

More generally, we consider context as everything that surrounds a utilisation situation of a knowledge system by a user and can throw light on its meaning. Therefore, context is characterised by interrelated conditions for the existence and occurrence of the utilisation situation such as the external environment, the internal state, location, time, history, etc. For knowledge systems we need to handle the mental context that is based on the profile of the actor or user, the storyboard context that is based on the story leading to a situation, the data context that is based on the available data, the stakeholder context, and the collaboration context. These different kinds of contexts have an influence on the development of the storyboard and must thus be considered for the development of the knowledge system.

We distinguish the following facets of context [34,33,30]:

Actor context: The knowledge system is used by actors for a number of tasks in a variety of involvements and well understood collaboration. These actors impose their quality requirements on the knowledge system usage as described by their security and privacy profiles. They need additional auxiliary data and auxiliary functions. The variability of use is restricted by the actor's context, which covers the actor's specific tasks and specific data and function demand, and by chosen involvement, while the profile of actors imposes exceptions. The involvement and collaboration of actors is based on assumptions of social behaviour and restrictions due to organisational decisions. These assumptions and restrictions are components of the actor's context.

Storyboard context: The meaning of content and functionality to users depends on the stories, which are based on scenarios that reflect life cases and the portfolios of users or actors. According to the profile of these users a number of quality requirements such as privacy, security and availability must be satisfied. The actor's scenario context describes what the actor needs to understand in order to efficiently and effectively solve his/her tasks in the actual portfolio. The actor's determine the policy for following particular stories.

System context: The knowledge system is developed to support a number of intentions. The purposes and intents lead to a number of decisions on the knowledge system architecture, the technical environment, and the implementation. The knowledge system architecture has an impact on its utilisation, which often is only implicit and thus leads to not understandable systems behaviour. The technical environment restricts the user due to restrictions imposed by server, channel and client properties. Adaptation to the current environment is defined as context adaptation to the current channel, to the client infrastructure and to the server load. At the same time a number of legal decisions based on regulations, laws and business rules have been incorporated into the knowledge system.

Temporal context: The utilisation of a scene by an actor depends on his/her history of utilisation. Actors may interrupt and resume their activities at any moment of time. As they may not be interested in repeating all previous actions they have already successfully completed, the temporal context must be taken into account. Due to availability of content and functionality the current utilisation may lead to a different story within the same scenario.

Provider context: Providers are characterised by their mission, intentions, and specific policies. Additionally, terms of business may be added. Vendors need to understand how to run the knowledge system economically. Typical parts of this context are intentions of the provider, themes of the website, mission or corporate identity of the site, and occasion and purpose of the visits of actors. Thus, providers may require additional content and functionality due to their mission and policy. They may apply their terms of business and may require a specific layout and playout.

Based on this information, the knowledge system is extended by provider-specific content and functionality. The storyboard may be altered according to the intentions of the provider, and life cases may be extended or partially supported. Provider-based changes to portfolios are typical for knowledge systems in e-government and e-business applications.

Developer context: The knowledge system implementation depends on the capability of the developer. Typically we need to take into account the potential environment, e.g. hard- and software, communication channels, the information systems that are to be incorporated, especially the associated databases, and the programming environment developers use.

Organisational and social context: The organisation of task solutions is often already predetermined by the application domain. It follows organisational structures within the institutions involved. We captured a part of these structures already on the basis of the portfolio and modelled it by collaboration. The other pars form the organisational context. Collaboration of partners consists of communication, coordination, and cooperation. Cooperation is based on cooperativity, i.e. the disposition to act in a way that is best helpful for the collaboration partners, taking their intentions, tasks, interests and abilities into account. At the same time, collaboration is established in order to achieve a common goal. Actors choose their actions and organise them such that their chances of success are optimised with respect to the portfolio they are engaged in. Additionally, the social context may be taken into account, which consists of interactive and reactive pressures. Typical social enhancements are socially indicated situations such as welcome greetings, thanking, apologising, and farewell greetings.

Most systems today do not support adaptivity and user orientation. Information as processed by humans is perceived in a very subjective way. As for a knowledge system, the determining factor whether the user can derive advantage from the content delivered is the user's individual situation, i.e. the life case, user model and context. The same category of information can cause various needs in different life cases.

Not any user can deal with any kind of content. For the casual user or the novice other content has to be delivered than for experts. The common knowledge system doesn't reflect the user's situation and neglects the user's specific needs. As a result, the user is spammed with information which is predominantly out of focus. The abundance of information also makes it impossible to separate useful from for the user useless content. Any by the absence of meta data unspecified information reduces the usability of World Wide Web on the whole.

Furthermore, users are limited

- in their abilities for verbalisation,
- in their abilities for digestion of data and
- by their habits, practices and cultural environment.

These limitations may cause intellectual overburdening of users. Most systems that require sophisticated learning courses for their exploration and utilization did not consider these limitations and did not cope with real life situations. The approach we use for avoiding overload is based on observation of real applications before developing the knowledge system.

User typically request or need various content depending on their situation, on material available, on the actual information demand, on data already currently available and on technical equipment and channels on hand. Therefore, we need a facility for content adaptation depending on the context of the user. Content matching and adaptation may be thus considered as one of the 'grand' challenges of modern internet.

To meet this challenge, the information has to be matched against the particular needs of the user [34,33,30]. Since the thinkable combinations of user life cases, user models and context [15] are indefinitely, the definition of life cases

[33] has to be determined for the content and matched against the users situation. For a knowledge system, there should be not only concrete definitions of which content is applicable for which life case. To avoid making useful content useless by presenting it in an inappropriate way to the user, knowledge systems have also to consider the user's specific profile and context. By processing this data, the knowledge system should provide different views of information and the appropriate media types for presenting their knowledge to various audiences.

The implicit goals of content management and content delivery are:

- to meet all the information (contextual) requirements of the entire spectrum of users in a given application area;
- to provide a "natural" and easy-to-understand structuring of the information content;
- to preserve the designers entire semantic information for a later redesign;
- to achieve all the processing requirements and also a high degree of efficiency in processing;
- to achieve logical independence of query and transaction formulation on this level;
- to provide a simple and easily to comprehend user interface family.

6. Conclusion

6.1. Knowledge in the Two Facets

The notion of knowledge is one of the most overused terms. It can be considered as knowledge in general defined by a noun (objective knowledge) and the knowledge by a user expressed by the verb 'to know' (subjective knowledge).

We conclude therefore that within the scope of the knowledge system environments, it is necessary to deliver knowledge as enduring, justified and true consensus to users depending on context, users demands, desiderata and intention, whereby these aspects are supported by social facets, the environment, the profile, tasks and life cases of the users. Life cases, portfolios and tasks constitute the information demand of every users. The information demand of users requires a special content quality. It results in the requested knowledge, which is also depending on the understanding and motivation of users. So, the requested knowledge of users is a composition of understanding, and information demand, whereby the information demand is an aggregated component of life cases, motivation, intention and quality.

It is surprising that the literature treats knowledge on a 100 % quality basis. We can however distinguish between

validated knowledge that satisfiable within a scope of axioms and derivation rules (application domain), within a certain generality and has validity and timelineness,

verified and correct knowledge based on axioms and rules within a proof system that can be verified within a finite time, obey a correctness criteria (depending on profiles) and has some known interaction with other knowledge and finally

quality knowledge defined by the quality of use (understandability, learnability, operability, attractiveness, appropriatedness), by the external quality (privacy, ubiquity, pervasiveness, analysability, changeability, stability, testability), and by the internal quality (accuracy, suitability, interoperability, robustness, self-contained/independence).

These quality characteristics result in differences of the value of knowledge for the user. Sections 2, 3 and 4 demonstrate the challenges we face with knowledge system environments. Our general notion of knowledge is supported by current database and information systems technology based on a separation into content, concepts, topics and memes in Section 5.

6.2. Knowledge Management Systems

Knowledge management aims in supporting the spread of knowledge of individuals or groups across interested in it communities in ways that directly affect performance. Knowledge management envisions getting the *right data* (or information) within the *right context* to the *right person* at the *right time* for the *right business purpose*. Knowledge management systems allow to manage knowledge in communities of practice or interest, supporting creation, capture, and storage and sharing of expertise in the form of information in dependence on the user. Knowledge management systems use the entire variety of architecture solution known for intranet, internet or extranet systems. We observe two main architecture lines:

1. The task/profile/portfolio-based approach focuses on the use of knowledge by collaborators depending on their tasks and interest. This approach is based on the data or information and knowledge needs of the communities, where they are located, and who needs them. The KMS is designed to capture knowledge and to make knowledge available when needed to whom needs it. These systems may be described by storyboards [18,31] that describe the life situations or life cases of the knowledge demanders or user, the context of knowledge use and the user profiles which specify the education, personality and practice profile.
2. The infrastructure/generic system based approach focuses on building a knowledge base system to capture and distribute knowledge for use throughout the communities of practice. It concern of the technical details needed to provide good mnemonic functions associated with the identification, retrieval, and use of knowledge. The approach focuses on network capacity, database structure and organization, and knowledge information classification.

Sections 3 and 4 describe one approach how such knowledge management systems can be developed. These approaches combine the two architecture lines. Section 5 develops a background for a general knowledge system environment.

6.3. Semantification of Web 3.0

The "Semantic Web" is mainly based on syntax and partially uses micro-semantics of wordings. Semantics is used in the sense of rudimentary lexical semantics.

Rudimentary lexical semantics must be enhanced by explicit definitions of symbols or words used. These definitions can be combined with the name spaces that provide a source for the lexical units used in a web document. The semantification project [6] of the group headed by J. Pokorny and P. Vojtas at Charles University Prague aims in enhancing the ontology based annotation in XML documents or RDFa-annotated HTML files by a semantic repository, by user profiles and by portfolio management.

Web documents should be enhanced by context [15] or meta-data similar to the OpenCyc project. Lexical units may be characterised by *time(absolut, type), place(absolut, type), culture, sophistification/security, topic/usage, granularity, modality/disposition/epistemology, argument preferences, justification,* and *lets* [Len02].

The vocabulary of name spaces or of ontologies is not just a collection of words scattered at random throughout web documents. It is at least partially structured, and at various levels. There are various modes and ways of structuring, e.g., by branching, taxonomic or meronymic hierarchies or by linear bipole, monopolar or sequenced structures.

Ontologies are often considered to be the silver bullet for web integration. They are sometimes considered to be an explicit specification of conceptualisation or to be a shard understanding of some domain of interest that can be communicated across people and computers. We should however distinguish a variety of ontologies such as generic, semiotic, intention/extension, language, UoD, representational, context and abstraction ontologies.

6.4. Technical Environments for Knowledge on Demand

In this paper we discussed three approaches to technical environments to knowledge systems: cloud services, database-backed systems and knowledge processing for universal communities. Technical environments for knowledge delivery system includes pull and push technology, notification technology, knowledge discovery technology, knowledge documentation, knowledge quality and productivity, and human computer interface technology. These systems be developed by using eight layers, which includes storyboard layer as a top level one in order to allow to provide knowledge on demand and on context, and followed by seven interwoven technology layers that facilitate the community to work together to share, re-use and generate knowledge among them: interface layer, access layer, collaborative layer, application logic layer, transport layer, integration layer, and repositories layer.

References

[1] S. Akamine, Y. Kato, D. Kawahara, K. Inui, S. Kurohashi, and Y Kidawara. Development of a Large-scale Web Crawler and Search Engine Infrastructure. Proceedings of the 3rd International Universal Communication Symposium (IUCS2009), 126-131, 2009

[2] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, Above The Clouds: A Berkeley View of Cloud Computing. *Technical Report No. UCB/EECS-2009-28*. EECS Department. University of California, Berkeley. 2009.

[3] R.Buyya, C. Shin Yeo, and S. Venugopal, 2008. Market-Oriented Cloud Computing: Vision, Hype and Reality for Delivering IT Services as Computing Utilities. *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications* (HPCC-08, IEEE CS Press, Los Alamitos, CA, USA), Sept. 25-27, 2008, Dalian, China. 9p. 2008.

[4] A. Dey, B. Kokinov, D. Leake, and R. Turner (Eds.). Modeling and using context. *5th International and Interdisciplinary Conference* CONTEXT 2005. Springer. Berlin, 2005.

[5] Dublin Core Metadata Initiative. Dublin Core. http://dublincore.org/, June 2007.

[6] M. Duží, A. Heimburger, T. Tokuda, P. Vojtas, and N. Yoshida. Multi-agent knowledge modelling. In H. Jaakkola and Y. Kiyoki, editors, *EJC'2008*, Information Modeling and Knowledge Bases XVI. IOS Press, 2008. Panel summary, EJC 2008.

[7] G. Fiedler and B. Thalheim. Towards semantic wikis: Modelling intensions, topics, and origin in content management systems. *Information Modelling and Knowledge Bases*, XX:1–21, 2009.

[8] X. Fu, T. Bultan, and J. Su. Analysis of Interacting BPEL Web Services. Proce. of the 13th international conference on *World Wide Web*, pp. 621–630, 2004.

[9] C. Ghidini and F. Giunchiglia. Local models semantics, or contextual reasoning = Locality + compatibility. http://citeseer.ist.psu.edu/481285.html, April 2000.

[10] V. Haarslev and R. Möller. Racer: An OWL reasoning agent for the semantic web. In *Proceedings of the International Workshop on Applications, Products and Services of Web-based Support Systems, in conjunction with the 2003 IEEE/WIC International Conference on Web Intelligence, Halifax, Canada, October 13*, pages 91–95, 2003.

[11] V. Haarslev, R. Möller, and M. Wessel. Description Logic inference technology: Lessions learned in the trenches. In I. Horrocks, U. Sattler, and F. Wolter, editors, *Proc. International Workshop on Description Logics*, 2005.

[12] T. Hey, S. Tansley, and K. Tolle. The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research (2009).

[13] Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen. From SHIQ and RDF to OWL: The making of a web ontology language. *Journal of Web Semantics*, 1(1):7–26, 2003.

[14] H. Jaakkola, A. Heimbürger, and P. Linna. Knowledge-oriented software engineering process in a multi-cultural context. *Software Quality Control*. Vol 18 (2). pp. 299-319, 2010.

[15] R. Kaschek, K.-D. Schewe, B. Thalheim, and Lei Zhang. Integrating context in conceptual modelling for web information systems, web services, e-business, and the semantic web. In *WES 2003*, LNCS 3095, pages 77–88. Springer, 2003.

[Len02] D. Lenat. The dimensions of the context space. www.cyc.com/context-space.pdf, 2002.

[16] J. Lewerenz, K.-D. Schewe, and B. Thalheim. Modeling data warehouses and OLAP applications by means of dialogue objects. In *Proc. ER'99, LNCS 1728*, pages 354–368. Springer, Berlin, 1999.

[17] R. Lewis. *The Cultural Imperative*. Global Trends in the 21st Century. Intercultural Press. Brealey. Printed in Finland. 338p., 2007.

[18] T. Moritz, K.-D. Schewe, and B. Thalheim. Strategic modelling of web information systems. *International Journal on Web Information Systems*, 1(4):77–94, 2005.

[19] G. L. Murphy. *The big book of concepts.* MIT Press, 2001.

[20] T. Nakanishi, H. Homma, K.-S. Kim, K. Zettsu K., Y. Kidawara, and Y. Kiyoki. A Three-layered Architecture for Event-centric Interconnections among Heterogeneous Data Repositories and its Application to Space Weather. Proc. of the 20th *European Japanese Conference on Information Modelling and Knowledge Bases*, EJC 2010, Jyväskylla, Finland, pp. 29–44, 2010.

[21] T. Nakanishi, K. Zettsu, , Y. Kidawara, and Y. Kiyoki. A Context Dependent Dynamic Interconnection Method of Heterogeneous Knowledge Bases by Interrelation Management Function, In Proc. of the 19th *European-Japanese conference on information modelling and knowledge bases*, EJC 2009, pp. 210-227, 2009.

[22] , T. Nakagawa, K. Inui, and S. Kurohashi. Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables. Proc. of *Human Language Technologies*: The 11th Annual Conference of the North American Chapter of the Association for Computational

Linguistics (HLT-NAACL 2010), 2010
[23] NGG: Next Generation Grids Expert Group. Future of European Grids: Grids and Service-Oriented Knowledge Utilities. Report 3, 2006.
[24] M.P. Papazoglou, and D. Georgakopoulos. Service-Oriented Computing. *Communications of the ACM* Vol. 46, No. 10, pp. 24–28, 2003.
[25] L. Reeve and H. Han. Survey of semantic annotation platforms. In *SAC '05*, pages 1634–1638, New York, NY, USA, 2005. ACM Press.
[26] M. Rönkkö, J. Ylitalo, J. Peltonen, N. Koivisto, O. Mutanen, J. Autere, A. Valtakoski, and Pentikäinen, National Software Industry Survey 2009. *Helsinki University of Technology*. Helsinki. 128 p., 2009
[27] J.E. Safra, I. Yeshua, and et. al. *Encyclopædia Britannica*. Merriam-Webster, 2007.
[28] K.-D. Schewe and B. Thalheim. Reasoning about web information systems using story algebra. In *ADBIS'2004*, LNCS 3255, pages 54–66, 2004.
[29] K.-D. Schewe and B. Thalheim. The co-design approach to web information systems development. *International Journal of Web Information Systems*, 1(1):5–14, March 2005.
[30] K.-D. Schewe and B. Thalheim. Usage-based storyboarding for web information systems. Technical Report 2006-13, Christian Albrechts University Kiel, Institute of Computer Science and Applied Mathematics, Kiel, 2006.
[31] K.-D. Schewe and B. Thalheim. Usage-based storyboarding for web information systems. Technical Report 2006-13, Christian Albrechts University Kiel, Institute of Computer Science and Applied Mathematics, Kiel, 2006.
[32] K.-D. Schewe and B. Thalheim. Development of collaboration frameworks for web information systems. In *20th Int. Joint Conf. on Artifical Intelligence, Section EMC07 (Evolutionary models of collaboration)*, pages 27–32, Hyderabad, 2007.
[33] K.-D. Schewe and B. Thalheim. Life cases: A kernel element for web information systems engineering. In *Web Information Systems and Technologies*, volume Volume 8. Lecture Notes in Business Information Processing, Springer Berlin Heidelberg, 2008.
[34] K.-D. Schewe and B. Thalheim. User models: A contribution to pragmatics of web information systems design. In K. Aberer, Z. Peng, and E. Rundensteiner, editors, *Web Information Systems – Proceedings WISE 2006*, volume 4255 of *LNCS*, pages 512–523. Springer-Verlag, 2006.
[35] K.-D. Schewe and B. Thalheim. Semantics in data and knowledge bases. In *SDKB 2008*, LNCS 4925, pages 1–25, Berlin, 2008. Springer.
[36] John F. Sowa. *Knowledge Representation, Logical, Philosophical, and Computational Foundations*. Brooks/Cole, a division of Thomson Learning, Pacific Grove, California, 2000.
[37] B. Thalheim. The conceptual framework to user-oriented content management. *Information Modelling and Knowledge Bases*, XVII:30–49, 2007.
[38] W3C. Web Ontology Language Overview. http://www.w3.org/TR/owl-features/, Feb 2004.
[39] W3C RDF Core Working Group. Resource Description Framework (RDF). http://www.w3.org/RDF/, 2004.
[40] L. Youseff, M. Butrico, and D. Da Silva. Toward a Unified Ontology of Cloud Computing. *Grid Computing Environments Workshop*, GCE '08, pp.1-10, 2008.
Youseff, L.; Butrico, M.; Da Silva, D.; , "Toward a Unified Ontology of Cloud Computing," Grid Computing Environments Workshop, 2008. GCE '08 , vol., no., pp.1-10, 12-16 Nov. 2008
[41] S. Xu and W. Zhang. Knowledge as a Service and Knowledge Breaching. *Proceedings of the 2005 IEEE International Conference on Services Computing*. Vol 01. pp. 87-94, 2005.
[42] K. Zettsu, T. Nakanishi, M. Iwazume, Y. Kidawara, and Y. Kiyoki. Knowledge Cluster Systems for Knowledge Sharing, Analysis and Delivery among Remote Sites. *Information Modeling and Knowledge Bases*, Vol. XIX, IOS Press, pp.282–289 (2008).
[43] K. Zettsu, Y. Kidawara, and Y. Kiyoki. Developing Next Generation Web as Collaboration Media. *Infocommunications Journal*, Vol. LXV., No. 2010/I, pp.15–19, Scientific Association for Infocommunications, Hungary, 2010.