THESEUS - tool for internet services

3 March 2009

(PhysOrg.com) -- The improved use and exploitation of digital knowledge - that is the aim of the THESEUS Project. In the future semantic technologies will be able to recognise the meaning of information content. Fraunhofer researchers will be presenting the initial results at CeBIT.

"The society of the future will be even more knowledge-based than the present one. For that reason it is not only necessary to create the appropriate infrastructure, but also to ensure that existing knowledge is suitably prepared and made recognisable", explains Prof. Hans-Joachim Grallert, Head of the Fraunhofer-Institute for Telecommunications, Heinrich Hertz Institute HHI. In a joint project, entitled THESEUS, 31 German companies, universities and research institutes have banded together with the aim of enabling the improved use and exploitation of digital knowledge. Nine Fraunhofer Institutes are taking part in this project. The HHI is coordinating the development of the base technologies for THESEUS. The focus here is on semantic technologies for the next generation of the internet, which will recognise the meaning of information content and be able to classify it - irrespective of whether it be words, photos, sounds, 2D and 3D image data. With these technologies, computer programmes will be able to intelligently understand in what context data should be stored, as well as draw logistical conclusions and establish correlations. CeBIT will see the companies and researchers present the first results of the project.

In libraries, broadcasting institutions, archives, museums and databases the wealth of knowledge of our society lies slumbering. But how to better utilise it as well as make it accessible to a wide audience? Researchers from the Fraunhofer Institutes for Intelligent Analysis and Information Systems IAIS are working on the digitalisation of media types such as text, images and sound recordings and to connect the data semantically - that is to say in accordance with its contextual meaning - to an innovative knowledge network. This will enable people looking for information to undertake searches more easily.

To achieve the best-possible digitalisation results, HHI researchers are developing algorithms for the restoration of text and video data. These automatically detect and rectify faulty data reliably. They can, for example, optimise yellowed pages for text recognition or clean historical film recordings of dust and scratches. New quality analysis processes for images and videos will in the future actually be able to automatically detect defects such as dust or scratches in images and videos or identify quality characteristics such as sharpness and contrast. This information serves as the basis for the digital restoration of the material or for searching qualitative high-grade content.

Up to now the searching of video and photo archives has been particularly time-consuming and laborious. In the future, metadata - a type of specification of content - for multimedia data will be automatically generated and thereby simplify the search. Taking images for example: researchers are developing image recognition systems that utilise colours or structures in the image - textures - to makes inferences regarding the content. This would enable a computer to identify, by way of example, sun, blue ocean or geometric forms such as beach chairs and to store the image content in metadata.

New possibilities are offered by the intelligent searching of images in medicine. In a sub-project called Medico, the development efforts of experts from the Fraunhofer Institute for Computer Graphics Research (IGD), include FIRST computer and software technology and IAIS tools for the automatic statistic evaluation of medical image data such as computed tomography images, for example. This will enable the future close matching of image characteristics to the symptoms of disease. The procedure will allow images of one patient to be compared at lightning speed with the images of thousands of other patients. This makes it easier for physicians to make a diagnosis. The IGD researchers are working on the functional

graphical interface of THESEUS. "The knowledge infrastructure of the future will enable the user to intuitively and quickly find and assess all the services required for a particular subject", explains Nadeem Bhatti, research assistant at the IGD. Using the graphical interface it is possible, for example, to identify all services on topics such as "Calculation of the eco-balance of products" or "medical diagnoses". The underlying data will retrieved and processed with the assistance of THESEUS technologies.

THESEUS:

THESEUS is a research program initiated by the Federal Ministry of Economics and Technology (BMWi) and is aimed at the development of an new internet-based knowledge infrastructure, for improving the utilisation and exploitation of knowledge via the internet. The focus of the research program is on semantic technologies that can identify content (words, images, sounds) not with the assistance of conventional processes (e.g. letter combinations), but instead that can recognise the meaning of the content of information and classify it. With this technology, computer programs can understand in what context data should be stored. Furthermore, by applying the rules and order principles, computers can draw logical conclusions about content and recognise and create independent correlations between various pieces of information from several sources.

Provided by Fraunhofer-Gesellschaft