

Towards Evaluating the Core Technology Cluster of the German Research Program THESEUS

Extended Abstract

Peter Dunker, Juan José Bosch Vicente, Judith Liebetrau
Fraunhofer Institute of Digital Media Technology, Ilmenau, Germany

THESEUS is a research program¹ initiated by the German Federal Ministry of Economy and Technology (BMWi), with the goal of developing a new Internet-based infrastructure in order to better use and utilize the knowledge available on the Internet. THESEUS consists of a core technology cluster (CTC) developing basic technologies, and application scenarios which package the core technologies and concentrate on the development of concrete use case applications. Fraunhofer Institute for Digital Media Technology (IDMT) is the lead of the evaluation task which is responsible for the quality assessment of the core technologies developed by several research partners involved in the CTC of the THESEUS program. Due to the wide spread fields of research of basic technologies in THESEUS, there are several related fields of quality assessment such as text recognition, image and video analysis, privacy and security aspects, watermarking, video and audio quality, iterative system design and quality in use and field testing. In this paper we will give an overview over the different evaluation tasks and approaches, which will be conducted by Fraunhofer IDMT.

Text Recognition

The handwriting text recognition technology will be evaluated with a set of hand written data of a determined domain, due to the fact that the recognition utilizes a priori knowledge of specific domains. The recognition rate of the algorithm is investigated for different distortion levels of the test data. Furthermore, a comparison with publicly available text recognition systems and handwriting databases is planned.

Image and Video Analysis

The different kind of image and video analysis approaches will be divided into major test cases e.g. retrieval and classification. Retrieval

refers to algorithms which return a list of various media items by any kind of query. Classification refers to algorithms which return a description of a specific media query. A generic evaluation framework and tool set will be developed in order to perform the assessment of these tasks. The key features of the framework are: easy extension to new formats and measures, storing previous test results for comparison and measurement of improvements, sophisticated visualizations for interactive reviewing and generation of descriptive test results. A more detailed description of the image and video analysis evaluation is provided in a separate paper at the ImageClef workshop.

Privacy and Security

With respect to privacy aspects, and certain security aspects within that context, the following tasks will be evaluated: Cross-Platform, interoperable Digital Rights Management; Watermarking; Conditional Access and Content Protection, Encryption; Labeling, Filtering and Identification.

The evaluation concentrates on the analysis of components / (sub)systems where sensitive information is communicated and/or processed. In order to identify possible privacy problems, a holistic approach is applied, including legal, technical, organizational and economical aspects. After analyzing the system and respective data flows and traces, the system will be evaluated based on the following criteria: Data minimization; Transparency; Multilateral Security; System Integration; User Empowerment. As a result of the evaluation, possible problems will be identified and, where applicable, proposals for technical and organizational improvements will be provided.

Watermarking

Digital watermarking is the process of embedding information into multimedia signals (audio, video and picture). These

¹ www.theseus-programm.de/

embedded signals should be invisible, inaudible and robust against manipulation of the supporting medium.

Via listening and visual tests, the influence of perceptual quality caused by embedding a watermark will be evaluated. The quality assessment will be conducted according to standardized test procedures like double blind tests [1].

With robustness test the resistance against changes in the multimedia signals caused by reproduction-, encoding- and transmitting process will be verified. The multimedia data will be processed according to established test procedures [2] and the detectability of the watermark will be checked. During the project progression these signal processing procedures will be adjusted and adopted to the special needs of the THESEUS project.

Audio and Video Coding

The subjective assessment of audio- and video coding technologies will be performed according to well-established standards of ITU-R or other organizations. Due to IDMT's availability of standard-conformant test rooms (ITU-R BS.1116 / EBU tech 3276), a high precision assessment can be guaranteed.

Iterative System Design and Quality in Use

This task deals with the evaluation of diverse technologies, and its main purpose is to ensure their quality by following an iterative system design [3]. The developments will be tested against established standards, the use case requirements, and compared to state of the art technologies. The main criteria for the evaluation will be the technical quality of the developments, the quality in use of the technologies, and the user/stakeholder satisfaction.

Different approaches will be required in each of the cases, due to the diversity of technologies to be evaluated. The ontology management tools (design, mapping, evolution and reasoning) will be evaluated following pragmatic criteria, and performance issues (speed, scalability, etc.). The semantic access components will be individually evaluated. However, a global end-to-end evaluation will provide more useful results, by applying a multimodal evaluation framework and several standards [4] [5]. The semantic visualization framework will be evaluated both internally and also in relationship to other tasks. Standards for Human-System Interaction [6],

Software Quality [5], etc. and the use case requirements will serve as a basis for this evaluation. The statistical machine learning components will be mostly evaluated by using a method-based evaluation, computing the appropriate metrics, and comparing them to the values from state-of-the-art technologies and requirement specifications. Further evaluation criteria concern scalability, response time, or throughput analysis.

Field Testing

A small task within the evaluation concentrates on the assessment of both the user experience and the handling of the human computer interface by the interaction of multiple core technologies, utilizing empirical user tests.

Conclusion

The evaluation process will provide feedback to the THESEUS CTC partners, to be used in the subsequent iteration of the design process. Additionally, this evaluation will contribute to the improvement of evaluation techniques, and standards development.

References

- [1] "Methods for the Subjective Assessment of small Impairments in Audio Systems including Multichannel Sound Systems", Recommendation ITU-R, BS.1116-1*, 1994-1997
- [2] BPN 058 – Report on testing of audio watermarking systems, EBU Project Group B/AIM (Audio In Multimedia), August 2003
- [3] ISO 13407, Human-centred design processes for interactive systems
- [4] ISO 9241-110, Ergonomics of human-system interaction — Part 110: Dialogue principles
- [5] ISO/IEC 9126, Software engineering -- Product quality
- [6] ISO 9241-11, Ergonomic requirements for office work with visual display terminals (VDTs) -Part 11: Guidance on usability