The Dutch Language Institute: The Central Hub for Dutch Language Resources

Abstract

In this presentation we will discuss the needs for a language community to have extensive language resources and a digital language infrastructure for the development of all types of smart computer applications. We expand on how an open resource digital language infrastructure is made available for the Low Countries through the Dutch Language Institute. Such an infrastructure facilitates the development of a large number of technological applications.

In 2011, META, the Multilingual Europe Technology Alliance published a number of white papers discussing the benefits offered by Language Technology and the actions that need to be taken to develop basic tools and data for each language depending on the factors such as the complexity of the respective language, the size of its community, and the existence of active research centres in this area. Language technology is used to develop smart software systems designed to handle human language and is therefore often referred to as “human language technology”. Human language technology (HLT) links language to various forms of knowledge. Main application areas of language technology are a.o. language checking, web search, speech interaction, and machine translation. A large number of smart computer applications rely heavily on speech and language data; we name a few: spelling correction, authoring support, computer assisted language learning, information retrieval, information extraction, text summarization, question answering, speech recognition, speech synthesis,..

So far, the situation of Dutch concerning language technology support gives rise to some cautious optimism. Supported by larger research programs in the past, there exists a language technology industry and research environment in the Low Countries. For standard Dutch, a number of technologies and resources exist, but far less than for English. The Metanet consortium stresses the need for continuous development of language technology resources and use them to drive forward research, innovation and development. The need for large amounts of data and the extreme complexity of language technology systems makes it vital to develop a new infrastructure and a more coherent research organisation to support greater sharing and cooperation. In this paper, we will elaborate upon the role of the Dutch Language Institute as a developer, keeper and distributor of sustainable language resources.