Author's response to reviews

Title: Fine-grained Information Extraction from German Transthoracic Echocardiography Reports

Authors:

Martin F Toepfer (martin.toepfer@uni-wuerzburg.de)
Hamo Corovic (hamocorovic@googlemail.com)
Georg Fette (georg.fette@uni-wuerzburg.de)
Peter Klügl (pkluegl@uni-wuerzburg.de)
Stefan Störk (Stoerk_S@klinik.uni-wuerzburg.de)
Frank Puppe (frank.puppe@uni-wuerzburg.de)

Version: 2 Date: 17 March 2015

Author's response to reviews:

Clinical trials, research, and applications rely on the availability of appropriate data. Frequently, requested information is only stated in free text and can solely be accessed with information extraction methods. Although machine learning and adaptive approaches gained attention of researchers for several decades, manually reviewed domain specific representations are still required to specify the desired output. Construction of such resources is costly. As a result, many languages and domains have not been targeted sufficiently by now.

In the article "Fine-grained Information Extraction from German Transthoracic Echocardiography Reports" we describe a system that enables to gather fine-grained structured representations of semi-structured and unstructured German transthoracic echocardiography reports. In particular, we provide an evaluation of the component, and we give details about the process and the tools that were used to build the application.

Echocardiography is a vital source of information for clinical trials on heart failure, hypertension, myocardial infarction and coronary artery disease - just to name a few. Integration of the extracted information into a clinical data warehouse will support research in different fields.

This study extends previous work of the authors that already addressed information extraction from German clinical documents in several aspects. At first, the newly developed terminology is more meaningful, in terms of matching standardized concepts. Several entries have been extended, merged, and mapped. An improved knowledge representation and an enhanced information extraction pipeline with a different algorithm have been applied. Among others, we added filtering, layout analysis and subsection recognition to the application. This article provides a more detailed evaluation with macro-averaged and document layout specific results.
The manuscript has not yet been published in any other journal. All authors have agreed to its contents.

Authors’ Email:
martin.toepfer@uni-wuerzburg.de
hamocorovic@googlemail.com
georg.fette@uni-wuerzburg.de
pkluegl@uni-wuerzburg.de
Stoerk_S@klinik.uni-wuerzburg.de
frank.puppe@uni-wuerzburg.de