Information extraction of eligibility criteria for trial enrolment support

M. S. MARSHALL, J. VAN SOEST, R. DRESENS, J. PAULISSEN, P. LAMBIN, A. DEKKER

Maastricht University Medical Centre+, Department of Radiation Oncology (MAASTRO) GROW School for Oncology and Developmental Biology, Maastricht, The Netherlands

Keywords. Clinical Decision Support, Information Extraction

Rapid and efficient identification of patients eligible for clinical trials is on the critical path to the effective treatment of patients in a radiation oncology clinical setting. However, the Trial Physician Assistants (TPAs) that perform this work must read through many unstructured documents that have been received from regional hospitals and scanned in. The TPA must be familiar with a large set of clinical terms and abbreviations, as well as the inclusion and exclusion criteria of all active trials making the identification process a time-consuming and tedious task.

We have developed an approach based initially on Lucene (lucene.apache.org/) that produces daily reports that help to guide our TPAs more quickly to the information necessary to evaluate the eligibility of patients for trials. TPAs are able to quickly navigate to the context in a relevant document using hyperlinks in the report. The reports are produced on the basis of search patterns (Lucene queries) that we have designed to find mentions of eligibility criteria specific to trials at our clinic. We use statistics gathered during validation of our search patterns to provide confidence measures, as well as decide on report contents. Part of our approach involves the creation of a set of command line components in Ruby which support a rapid development and validation cycle for queries and flexible report generation. We are able to store information in both JSON files and RDF (www.w3.org/RDF/).

We now have a light-weight system consisting of an incremental indexer, report builder, query validation process, and statistics about each query that can automatically produce daily reports from lists of incoming patients.

We do not aim to replace the human expert in the eligibility evaluation process but to extend the reach of the TPA by reducing the cognitive load. We perform supervised learning by adapting the logic and queries when the TPA disagrees with the tool’s suggestion.

m.scott.marshall@maastro.nl