An Automata Based Approach to Biomedical Named Entity Identification

James Dowdall\textsuperscript{1}, Bill Keller\textsuperscript{1}, Lluís Padró\textsuperscript{2}, and Muntsa Padró\textsuperscript{2}

\textsuperscript{1} School of Cognitive and Computing Sciences
The University of Sussex
\{j.m.dowdall,bill\}@sussex.ac.uk
\textsuperscript{2} TALP Research Center
Universitat Politècnica de Catalunya
\{padro,mpadro\}@lsi.upc.edu

In this work, we present a new approach to Biomedical NE identification using an automata learning algorithm: Causal-State Splitting Reconstruction \cite{1}. This algorithm has previously been applied to Named Entity Recognition \cite{2} obtaining good results given the simplicity of the approach.

The same approach has been applied to Biomedical NE identification, using GENIA corpus 3.0, with 10-fold cross-validation. Our system attained $F_1 = 73.14\%$.

These results can be compared directly to \cite{3} and \cite{4}, which used the same data. First system obtains $F_1 = 57.4\%$ using ME Models, and the second one reports $F_1 = 79.2\%$ using SVMs. Both improve their results using post-processing techniques, reaching $F_1 = 76.9\%$ and $F_1 = 79.9\%$ respectively.

Our system does not use any post-processing techniques, and takes into account few features, so the results are considered very promising. In future work some post-processing will be developed to improve the results.

References