

# Authoring case based training by document data extraction

Christian Betz

Univ. of Würzburg, Chair for Artificial Intelligence and Applied Informatics  
email: [betz@informatik.uni-wuerzburg.de](mailto:betz@informatik.uni-wuerzburg.de)

Modeling is the bottleneck to successful implementation of information systems. In this paper, we propose an evolutionary approach to modeling based upon word processing documents. We applied our approach and software system to authoring of medical case based training systems.

Case based training provides a promising way of educating medical students and physicians. Various systems were developed and are in use, e.g. in the CasePort-Portal<sup>1</sup> for medical training cases. We developed D3Trainer and d3web.Train<sup>2</sup>, and independent authors wrote several systems for these, e.g. in Rheumatology, Gastroenterology, and Cytology.

However, authors need to either hand-code the content (usually as HTML) or use highly sophisticated authoring systems which require instructions and experience to master the complex systems. E.g. our authors used the knowledge modeling environment KnowME.

Case-based training relies on a large number of training cases, and teachers usually want to use their own content. Thus, time to become acquainted and effort necessary to author new cases are to be minimized to widen the bottleneck.

We came up with an evolutionary approach of modeling medical training cases where author enrich existing documents. Thus, they can use well-known tools (MS Word or Open Office), build upon material available already. While the first drastically flattens the learning curve, the latter reduces time-to-class on each case.

With our approach we carry further the ideas Felciano and Dev put into practice in their system Short Rounds<sup>3</sup>. They only presented pre-existing documents as an electronic patient record. For d3web.Train, we not only include textual and image findings, but also allowed the complete diagnostic cycle (order examinations, interpret results, state diagnoses, order treatment, and check performance) and feedback to the learners within the document.

Authors are to annotate their documents with the knowledge required: Bold-facing and underlining diagnoses controls their weight in feedback. Intro and Comments are to be added or marked. Relationships from findings to diagnoses are color-coded. The “grammar” of available markup is easy to learn and the parsing tool provides explicit feedback on how to improve documents.

As first evaluations show, our approach drastically reduces learning time from “several weeks to month” down to 10 minutes. Instead of days, it takes our authors only hours to complete a case. Promising, as these results are, we move on to a wider area of applications. The document parsing engine (Phoenix) is available open source<sup>4</sup>. It is also used for extracting law knowledge for an eLearning system and to model knowledge bases for the expert system d3web/Assist<sup>5</sup>.

## References

<sup>1</sup> <http://www.caseport.de>

<sup>2</sup> <http://www.d3webtrain.de>

<sup>3</sup> Felciano, R.M. , Dev. P:Multimedia Clinical Simulation based on Patient Records: Authoring, User Interface, Pedagogy. Technical Report, Stanford University School of Medicine, 1994, [http://smi-web.stanford.edu/pubs/SMI\\_Reports/SMI-94-0535.pdf](http://smi-web.stanford.edu/pubs/SMI_Reports/SMI-94-0535.pdf)

[4 https://sourceforge.net/projects/phoenix-ie/](https://sourceforge.net/projects/phoenix-ie/)

[5 http://www.d3web.de](http://www.d3web.de)

[6 http://www.knowit-software.de](http://www.knowit-software.de)