AUTOMATING COURSE ASSEMBLY

Onsophic Inc. is a Silicon Valley based startup focused on transforming learning. We are designing and building a modern, Web-based, educational platform that enables organizations to deliver online learning materials and classes according to any instructional strategy (e.g. authentic learning). The platform continuously gathers and analyzes learning data and can therefore be seen as the equivalent of Google Analytics for learning by providing organizations with a toolset to measure, analyse and discover what works and, more importantly, what doesn't work in training.

Corporate product training is one of Onsophic's main use cases. By improving product knowledge through personalized learning, companies will have more effective staff, higher producing partners and more satisfied customers. Apart from training their staff in a more effective way, the Onsophic platform will also allow companies to improve and enrich their learning content.

One of the first barriers a company experiences when moving from a file-based training approach to an e-learning environment is to import their existing content into the e-learning platform. For example, user manuals of the various products a company offers are only available through a set of PDF files. Often, there is a mix available of PDF files, online HTML files, online or offline slides, etc. The goal of this project is to improve this import process and find automated ways to transform raw learning materials into structured courses, i.e. automating the course assembly.

Automation of the course assembly process involves a number of different steps:

- parse the input source;
- detect learning modules and their learning activities (e.g. based on the headings);
- tag the detected modules and activities based on named entity recognition;
- estimate the Bloom level[1] for each detected activity.

In order to keep this project feasible, the scope is narrowed based on the following assumptions:

- only consider typical product user manuals as input (examples will be provided)
- the input format is HTML
- the language of the user manual is English

The expected skill set for this project:

- Java
- Web APIs

[1] https://en.wikipedia.org/wiki/Bloom%27s_taxonomy