Software documentation

There are three elements to the software (available at github.com/HTrustProject/SemanFinal2015):

Front end:

Software in folder **/frontend**
The frontend is a web application using Ruby on Rails.
It accesses a Lucene index for the semantic search and the backend for the search disambiguation.

Back end:

The backend creates the knowledge network (CiC network) based on Wikipedia data.

1. Research software (higher quality) in **/backend**
2. Alternative scalable solution (lower quality): at
   https://github.com/dnmilne/wikipediaminer/

It provides the font end with semantic information (used for user-guided disambiguation) and the indexer with semantic information from automatic disambiguation.

Indexer:

Software in folder **/indexer**
The indexer is a stand-alone java application.
It takes documents in pair tree format and creates a Lucene index.

Additional information

- The software contains an INSTALL.sh – the script produces a description of how to install and run the software.
- The knowledge network is not included (as it is way too large) but rather this is created by the backend software.
- A software demo is available at: *www.youtube.com/watch?v=2LiW_4X_6iU*

June 2015