Master Thesis Plan **Master Artificial Intelligence**

Date	:	**20th September, 2016**
Name and student number	:	**Stefanie Deckers, i6099223**
Name of the thesis examiner	:	**Kurt Driessens**
Name of the daily supervisor (if different from the thesis examiner)	:	
Title of the master thesis	:	**Development of a Recommender System for multiple online shops**
Expected final date	:	**25th January, 2016**

Signature thesis examiner **Signature master student**

………………………………………….. ………………………………………..

Assessment
The assessment will be based on the contents and form of the thesis, the presentation of this thesis and any product and/or software that accompanies the thesis.

Research plan

Description of the assignment:

The master thesis will be performed at the e-business company Braintags GmbH in Willich, Germany. They are hosting various online shops for different customers and want a universal recommender system for all shops. This recommender system will be developed in the scope of the master thesis.

The recommender system is supposed to work as an independent software component, such that every online shop Braintags is hosting can request recommendations from it. It will be implemented in Java using the tool-kit vert.x *(http://vertx.io/)*.

At the beginning of the research the available data for the different shops is going to be investigated to find an appropriate model that can serve every online shop. It is known yet that there is no explicit user feedback in terms of ratings or reviews. So the recommender system has to rely on implicit feedback like purchases, clicks, browsing behaviour etc. Some research has to be done to find a way how this information can be used to achieve suitable recommendations.

Next some effort will be made for finding the most appropriate recommendation technique for this setting. Recommendations techniques basically can be divided into collaborative, content-based, knowledge-based, demographic, community-based or hybrid recommendation techniques as described in *(F. Ricci, L. Rokach and B. Shapira: Introduction to Recommender Systems Handbook. In: Recommender Systems Handbook, pp. 1–35. Springer New York/Dordrecht/Heidelberg/London (2011)):*

- **Content-based** techniques make recommendations based on similarities between items, i.e. by finding items that are similar to the items that the user liked in the past.
- **Collaborative** approaches rely on preferences of similar users, where user similarities are calculated by similar purchases/clicks/ratings in the past. There are two different types of collaborative filtering approaches. Memory-based, also called neighbour-based, approaches find neighbours of a user or an item at runtime and use them to make recommendations. Model-based techniques learn an alternative model in advance by transforming users and items into the latent factor space, e.g. by singular value decomposition.
- **Knowledge-based** approaches require domain knowledge which is used to make recommendations. Such domain knowledge needs to be built and maintained by a knowledge-engineer and contains information about which items or item features match a user's needs or preferences.
- **Demographic** recommenders make use of any demographic information of the users like their age or language.
- **Community-based** recommenders require information from any social network and use the preferences of the user's friends to make recommendations.

The only techniques that come into question are collaborative filtering, content-based and demographic techniques. A knowledge-based approach cannot be implemented, because there are no domain experts that could build a knowledge-base for every online shop. Community-based recommenders require any kind of social network, that is not available within online shops. Collaborative filtering approaches only need feedback for different items of any users in the system. Moreover there is content-based information for items as well as demographic information for users available, which allows the use of content-based and demographic recommenders, respectively.

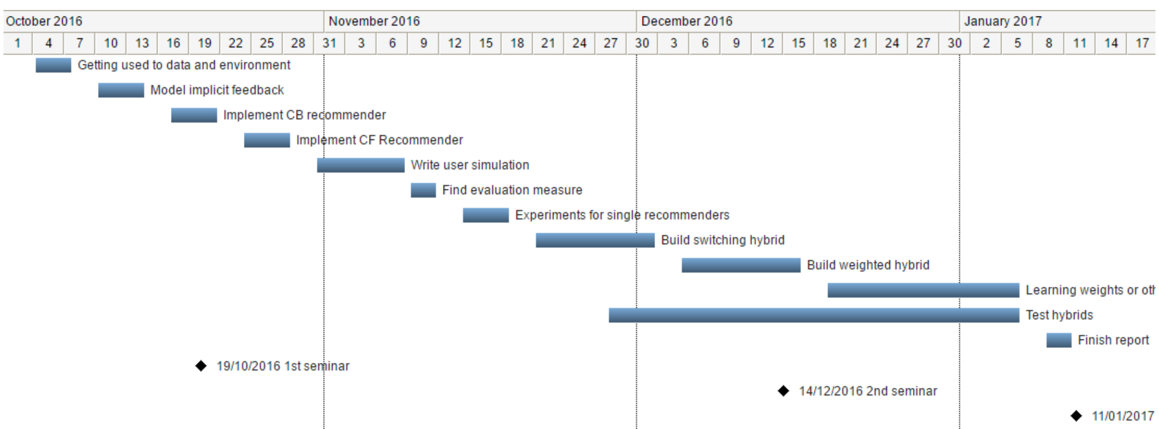
The goal of the master thesis is to develop a recommender system that produces sufficient recommendations for every online shop. The success of a recommendation techniques highly depends on the domain and the available data. For example collaborative filtering relies on purchases of other users: if there are not enough users with similar purchases as the current user, no confident recommendations can be made. Content-based recommenders need reliable information about items to determine a user's preferences and recommend appropriate items. Since the available data differs between different shops a hybrid recommender system probably is a good choice to benefit from different data sources (i.e. other users' ratings and content-based information). Different techniques as explained in (*Burke, R.: Hybrid web recommender systems. In: The Adaptive Web, pp. 377–408. Springer Berlin / Heidelberg (2007)*) will be implemented and tested to combine collaborative and content-based recommenders. If it is found that performance strongly varies among different online shops, a recommender will be developed that adapts to the online shops. One possibility is to implement a switching hybrid, that uses a different recommender system based on the current user profile. Another idea is to use a weighted hybrid, where recommendation is based on a score that is calculated as a linear combination from multiple recommenders. This sort of hybrid will be made adaptive by automatically learning the weights of that linear combination for every online shop.

To find a suitable recommender for the online shops, different recommenders will be implemented and evaluated. Evaluation will mainly be based on prediction accuracy using historical data. User behaviour will be simulated by starting at a specific past time point and hiding any future information. Recommendations will be generated using all available information up to that time point and compared to future information. For that purpose a measure has to be introduced to assess whether the recommendation was useful for the user, e.g. how many of the recommended items will be purchased by the user at future time points or "average rank of the correct recommendation" (ARC) as used in (*Burke, R.: Hybrid web recommender systems. In: The Adaptive Web, pp. 377–408. Springer Berlin / Heidelberg (2007)*).

Planning:

Name	Start	Finish
Getting used to data and environment • Investigate which data is available and how it can be assessed • Investigate how much is there for different shops and how is it distributed	04/10/2016	07/10/2016
Model implicit feedback • Research and write model	10/10/2016	14/10/2016
Implement CB recommender • Find item representation depending on available data • Find and implement recommender suitable for that representation (probably k-nearest-neighbour or Naive Bayes) • Add demographic information to content-based user profile	17/10/2016	21/10/2016
Implement CF recommender • Neighbourhood-based	24/10/2016	28/10/2016
Write user simulation for evaluation • Split into training and test set at different time points and run recommendations • sampling	31/10/2016	08/11/2016
Find evaluation measure • Research • Try e.g. accuracy or ARC and compare results	09/11/2016	11/11/2016
Experiments for single recommenders • Test both recommenders for different online shops • Try different parameters • Results will serve as baseline for later hybrid experiments	14/11/2016	18/11/2016
Build switching hybrid • Find decision criteria	21/11/2016	02/12/2016

Build weighted hybrid • Try some fixed weights	05/12/2016	16/12/2016
Learning weights or other improvements • If different weights work better for some shops, try to learn them. • Else try other hybrid, e.g. combined features	19/12/2016	06/01/2017
Test hybrids • Compare results to single recommenders	28/11/2016	06/01/2017
Finish report	09/01/2017	11/01/2017
1st seminar	19/10/2016	
2nd seminar	14/12/2016	
Handing in approval version	11/01/2017	
Handing in final version	25/01/2017	
Thesis presentation	31/01/2017	



In case of an external training period

Name of the company	:	Braintags GmbH
External supervisor	:	Martin Plücker
Size of the project	:	
Working hours	:	40h / week
Frequency of contact	:	Daily
Agreement about payment (if applicable)	:	400€ gross / month
Confidentiality	:	Confidentiality agreement regarding used data
Special agreements	:	

Signature external supervisor

…………………………………………………..