

Perceptum Project: Document recommendation for StarFish



Jorn Peters, Robbert van
Ginkel & Lotte Weerts

University of Amsterdam

June 5, 2014

Contents of this presentation

- Vision
- Product Backlog overview
- First sprint description
- Feedback



Product vision

About StarFish

- **Overall StarFish Vision:** People over all around the world need to get access to a network of knowledge in the Learning domain. It must be easy to browse the network and add new documents.
- **Overall StarFish Need:** This requires a network without a central authority: nobody knows all documents, so nobody can determine an entire structure.
- **StarFish Solution:** The StarFish website, containing all sorts of documents accessible for users. In the back-end, these documents are related in a graph.



Product vision

About our project

- **Problem:** currently, all users have to relate documents themselves. However, a user has no knowledge about all documents that are in StarFish that might be related.
- **Solution:** The StarFish system automatically recommends documents that are relevant to the users document.



Product vision

Elevator pitch

For StartFish users **who** search for and edit knowledge in Starfish, **the** document linker **is a** core system addition to StarFish **that** finds related documents. **Unlike** moderated or individual/centralized linking our product uses algorithms and data to automatically suggest document links.



Product Backlog

VALUE	POINT	WHO	WHAT	WHY	TYPE
1000	8	As a Perceptum employee	I want... to use an automated evaluation metric	in order to determine how well each algorithm performs at an early stage	Research/Implementation
800	8	As a Developer	I want... the system to use NN to calculate several distance metrics between document descriptors	in order to ensure the quality of the descriptor-based algorithms	Implementation
700	5	As a Perceptum employee	I want... to use the algorithm that uses document descriptors calculated with glossaries of tags	in order to see how it performs on the StarFish data set and compare it with other algorithms	Research/Implementation
600	2	As a Perceptum employee	I want... to use the algorithm that uses document descriptors based on the texts in documents	in order to see how it performs on the StarFish data set and compare it with other algorithms	Implementation
600	15	As a Perceptum employee	I want... to use the algorithm that uses document descriptors calculated using weighted tags in graphs	in order to see how it performs on the StarFish data set and compare it with other algorithms	Research/Implementation
600	8	As a Perceptum employee	I want... to use the algorithm that uses document descriptors calculated with weighted document descriptors based on text in documents	in order to see how it performs on the StarFish data set and compare it with other algorithms	Research/Implementation
500	2	As a Perceptum employee	I want... an evaluation metric based on expert examination	in order to find out in a detailed manner how well each of the algorithms performs	Research/Report/Implementation
400	?	As a Perceptum employee	I want... to know if there are other possible valuable algorithms	in order to find out in a detailed manner how well each of the algorithms performs	Research
300	100	As a Perceptum employee	I want... the implemented algorithms to be tuned towards the Starfish data	in order to find the best algorithm that gives the highest quality of the final product	Analysis/Implementation
50	20	As a Perceptum employee	I want... to have a clear overview (report) of all tried algorithms	in order to reuse the system and use it in the real Starfish application	Report

Product Backlog

Definition of Done

- The code is written OOP, and the classes that act correctly within the application framework
- The code is documented in a notebook containing examples of usage, test results and description of design choices
- The code is well tested both individually as within the application flow (this may be done manually no automated tests)



Sprint 1

POINTS	WHAT	TYPE
8	An automated evaluation metric	Research/Implementation
8	Implementation of NN to calculate several distance metrics between document descriptors	Implementation
5	Algorithm that uses document descriptors calculated with glossaries of tags	Research/Implementation
2	Algorithm that uses document descriptors based on the texts in documents	Implementation
15	Algorithm that uses document descriptors calculated using weighted tags in graphs	Research/Implementation
8	Algorithm that uses document descriptors calculated with weighted document descriptors based on text in documents	Research/Implementation
?	Find out if there are other possible valuable algorithms	Research



Feedback

We would love to hear what you think about our planning! Are there any suggestions?

