

# Retrospective Document

## Sprint 1

### Work & Test Progress

#### **Milestone 1) Instant Categorizer**

*Handling the case where a given webpage can't be found on database and needs to be categorized immediately.*

*(getUrl>checkDatabase>getContent>categorize>updateDatabase>returnCategory)*

*With 90% we have completed this milestone. We have created a java class named JSOUP, in order to get webpage content directly. Then we passed the content of it to the text to ARFF converter. Then we applied machine learning algorithm j48 in order to categorize on WEKA. After that we updated database.*

#### **Milestone 2) Expanding Category List**

*We have expanded number of categories from 2 to 10.*

*With 100% we have completed this milestone. We collect 100 URLs for each new category. In order to get supervised data, we have chosen URLs by hand searching that category name in Google. Then we choose more proper web sites which contains the most related textual content. We tried not to take from the same websites, therefore we branched to different web pages.*

#### **Milestone 3) Graphical User Interfaces**

*A GUI which integrates all the system.*

*With 100% we have completed this milestone. We have created a GUI which integrates all the system. Therefore, by using this interface, we can see results clearly on the screen.*

#### **TESTS:**

- *Use graphical interface and test button functionalities.*
- *Give a URL which is stored in database.*
- *Click search button.*
- *Check whether the category is fetched directly.*
- *Check whether the category of URL is returned correctly.*
- *Give a URL which is not stored in database.*
- *Click search button.*
- *Check whether the URL content is fetched from the web correctly.*
- *Check whether the content of it is converted to ARFF format correctly.*
- *Check whether categorization algorithm is applied correctly.*
- *Check whether the URL is categorized and showed in graphical interface correctly.*
- *Check whether the URL's category is added to database correctly.*

## Team Progress

- Onur Ozan Yüksel 25%
- İzzet Barış Öztürk 25%
- Mert Basmacı 25%
- Özge Donmaz 25%

## Left-overs (Backlog)

*We wanted to finish a working classifier as a milestone. We finished it but needs to be trained for every classification. We tried to save a model to use it for next classification queries but there is a little problem with saving the model. That problem makes the saved model unusable. Thus that milestone 1 is not finished 100%.*

## Next Sprint

**Milestone 1:** Training Set for 50 Categories Expanding the training datasets to cover 50 categories.

**Milestone 2:** Training Set for 100 Categories Expanding the training datasets to cover 100 categories.

**Milestone 3:** Training Set for ALL Categories We will try to cover the rest of the categories that our final product is required to. (Expanding the training datasets to cover 141 categories)

**Milestone 4:** Categorization Algorithms Trying out diferrent algorithms and finding the best categorization algorithm possibble for our project.

## Comments

*We had some issues due to incomplete parts of the product but we managed to solve the problem integrating the whole system. Now we have the skeleton of the project thus individual work will not be a problem. Our progress in the next sprint will be faster than ever.*

<WebCat Project>

<Overcode>

--

### Assistant's Evaluation

<i>Assistant's (Team Leader's) comments regarding to this completed sprint.</i>
---

### Supervisors's Evaluation

<i>Supervisor's (Team Leader's) comments regarding to this completed sprint.</i>
--