METU CENG491 2015 FALL

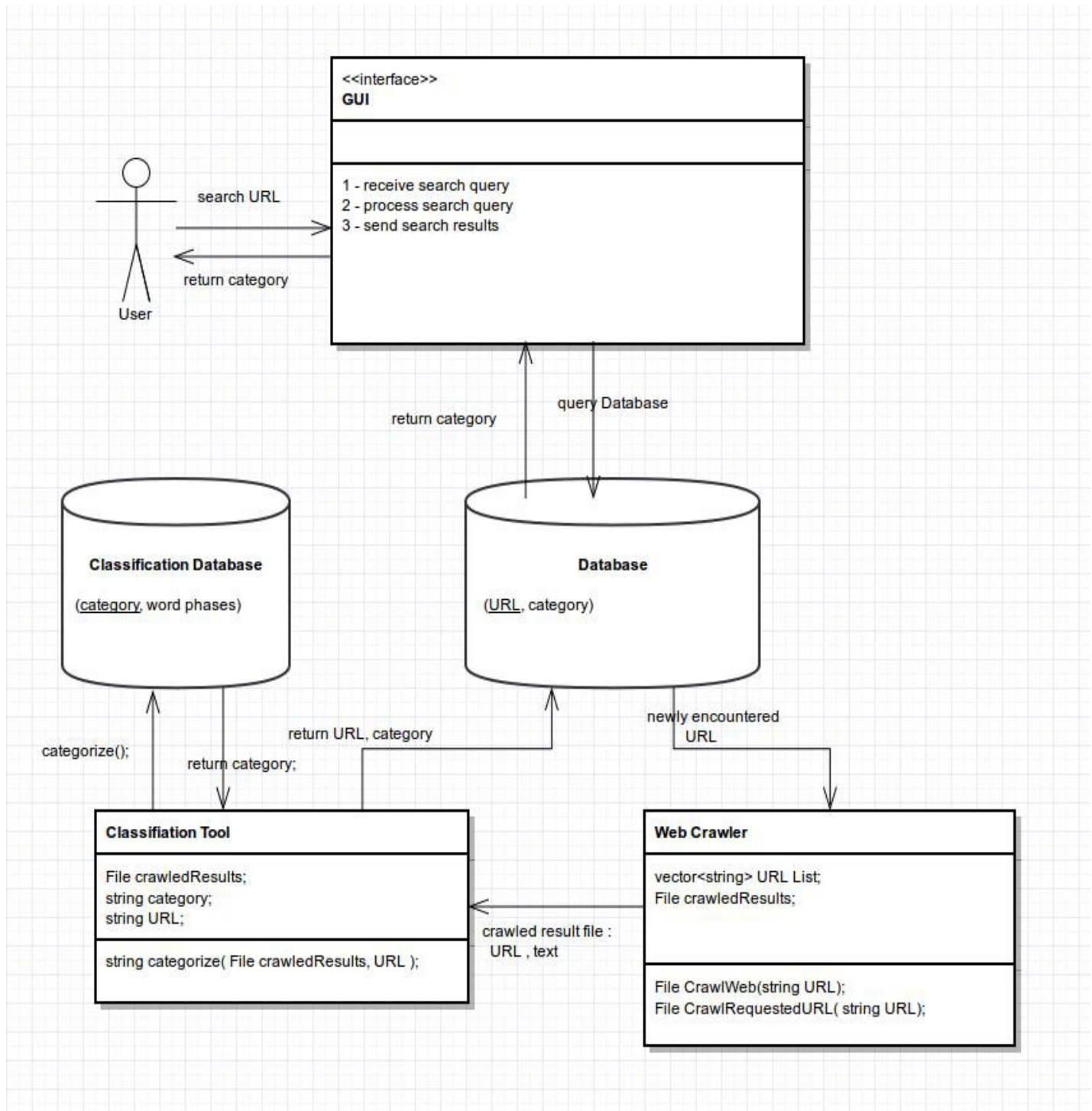## *START-UP DOCUMENT*

G05P43

Group Name: Overcode

Project Name: Web Page Categorizer

1. System Architecture

▪ _Draw_ the overall system architecture diagram. This should include (but it is not limited to) the components of the system, the interactions among the components and their dependencies.

▪ Identify and describe each component (including subcomponents if any), their _interactions_ and _dependencies_ clearly.

▪ Specify the user interaction model.

2. Tentative Time Plan

- *Identify and itemize <u>all tasks</u> to be performed as a team in the <u>first semester</u>. Assign a unique TaskID for each task. Give a short name and brief description for each identified task.*

TaskID	Short Name	Description
T1	Clarify	Clarifying the scope of the project, expectations and the desired outcome
T2	ML Research	Researching about Machine Learning Algorithms and Libraries to be used in the project.
T3	Database	Creating a database that meets requirements in a desired environment
T4	Samples	Creating Enough sample web pages to train the Machine Learning Algorithm
T5	Crawler	A crawler to get data for categorization process
T6	Training	Training the categorizing program via Machine Learning
T7	Categorizer	A fully working categorizing program that uses data that uses crawler
T8	Interface	An interface for final user usage
T9	Optimization	Crawling and Categorization optimization that enables the project to meet the 15 second limitation.

- *Construct your time plan as a simplified Gantt chart, as shown in the following table.*

	Iteration1	Iteration2	Iteration3
T1	■		
T2	■		
T3	■	■	
T4		■	
T5		■	■
T6		■	■
T7			■
T8			■
T9			■

3. Deliverables

- *Identify and list all deliverables of your project for the first 3 sprints.*

- *A deliverable is some component or sub-component, which is running and demonstrable to your assistant and your supervisor. That deliverable is of course subject to improvement over time.*

- *Fill in the following table:*

Deliverable	Description	When? (Sprint#)
D1	Crawler: A non stop working web site crawler	#2
D2	Categorizer: A program that categorizes a web page and updates the database	#3
D3	Database: A proper database that satisfies the project's needs	#1
D4	Interface: An interface for final user	#3

4. Workload Distribution

Fill in the following table to distribute the workload for the first semester among your team members.

	Sprint - I	Sprint - II	Sprint - III
Onur Ozan Yüksel	T1, T2	T4, T8 , T5	T5, T8
Barış Öztürk	T1,T3	T3, T4	T5, T8, T9
Mert Basmacı	T1, T3	T3, T4	T7, T9
Özge Donmaz	T1, T2	T4, T5, T6	T5, T6, T7