Software Requirements Specification

for

Wikipedia Mapper

**Version 2.0**

**Prepared by**

**Group Name: *Team WikiMapper***

|  |  |  |
| --- | --- | --- |
| **Ashley Revlett** | **1300073** | **anrevl01@louisville.edu** |
| **Everett Rush** | **1416168** | **enrush01@louisville.edu** |
| **Sarah Mullins** | **1853593** | **skmull02@louisville.edu** |

|  |  |
| --- | --- |
| **Instructor:** | **Dr. Ragade** |
| **Course:** | **CECS 550** |
| **Date:** | **4/24/2014** |

**Contents**

**Revisions**

**1** **Introduction**

1.1 Document Purpose

1.2 Product Scope

1.3 Intended Audience and Document Overview

1.4 Definitions, Acronyms and Abbreviations

1.5 References and Acknowledgments

**2** **Overall Description**

2.1 Product Perspective

2.2 Product Functionality

2.3 Users and Characteristics

2.4 Operating Environment

2.5 Design and Implementation Constraints

2.6 User Documentation

2.7 Assumptions and Dependencies

**3** **Specific Requirements**

3.1 External Interface Requirements

3.2 Functional Requirements

3.3 Behavior Requirements

**4** **Other Non-functional Requirements**

4.1 Performance Requirements

4.2 Safety and Security Requirements

4.3 Software Quality Attributes

**Revisions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Primary Author(s)** | **Description of Version** | **Date Completed** |
| 1.0 | Ashley Revlett, Sarah Mullins, Everett Rush | Initial version | 01/21/2014 |
| 2.0 | Sarah Mullins | Major revisions reflecting final status of project | 04/24/2014 |

# **Introduction**

Our project is to create software that will allow a user to search and explore Wikipedia in a visually-driven, graph-based way that clearly illustrates the connections between related articles. The requirements below outline in detail the way this software will work.

## Document Purpose

The purpose of this document is to outline the required functionality and design specifications for the software “Wikipedia Mapper,” version 2.0. This document outlines all components of the application, and is the only SRS for this software.

## Product Scope

Online media and web users are increasingly focused on visual multimedia content, as opposed to traditional text-based media. By providing user with a more intuitive, visually-driven way to explore the contents of Wikipedia, we can provide a more enjoyable, interactive way to learn from Wikipedia. Also, by visually showing the relationships between related topics, a user can get a quick overview of related topics, which can lead to further exploration of a topic. Also, by showing the interconnections between topics, users can more easily understand the “big picture” that surround a topic.

## Intended Audience and Document Overview

This document is intended for developers, designers, testers and the professor. It contains an overview of our program’s goals and features, and also outlines the technical requirements necessary for the software’s creation. Non-technical readers (designers, others) of this document will be most interested in section 2, which contains an overall description of the program. Technical readers (developers) will be most interested in section 3, which contains detailed implementation and design guidelines.

## Definitions, Acronyms and Abbreviations

*Active Article*: Wikipedia article currently being viewed.

*Node*: Object representing a Wikipedia article

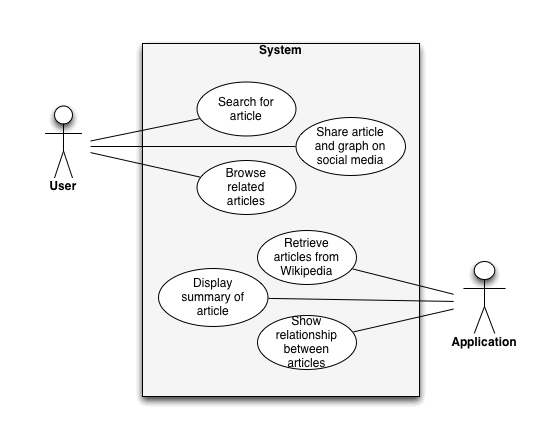
## References and Acknowledgments

[1] Wikipedia contributors. (2013, December 20). Wikipedia: Version 1.0 Editorial Team/Assessment [Online]. Available: http://en.wikipedia.org/wiki/Wikipedia:GRADE

# **Overall Description**

## Product Perspective

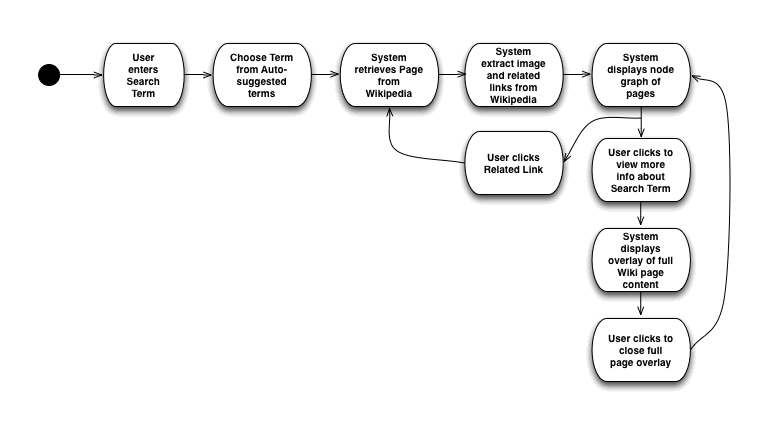
This is an entirely new product, although similar products have been produced by other companies in the past. Essentially, users will search for a subject using the software, and the software will retrieve the content from Wikipedia and visualize it in a novel way, using a node-based graph.



## Product Functionality

**Major Functions:**

* Display initial GUI with search box
* Retrieve page from Wikipedia
* Parse page to extract featured image and related links
* Display node graph of pages, with images
* Display overlay page with Wikipedia article summary



## Users and Characteristics

1. **Researchers and Students** – users who are looking for research material, and are interested in understanding the connections between topics. Frequent Wikipedia users who often prefer browsing images over text, but still need the detailed information text provides. Highly educated and comfortable with technology.
2. **Wikipedia Enthusiasts**– users familiar with Wikipedia, who are looking for a more interesting way to explore its contents, and often spend time exploring related topics in Wikipedia on their own. Highly educated and comfortable with technology.
3. **Casual Wikipedia users** – users who use Wikipedia on an as-needed basis, who are most focused on getting information quickly. Various education levels; may or may not be comfortable with technology.

## Operating Environment

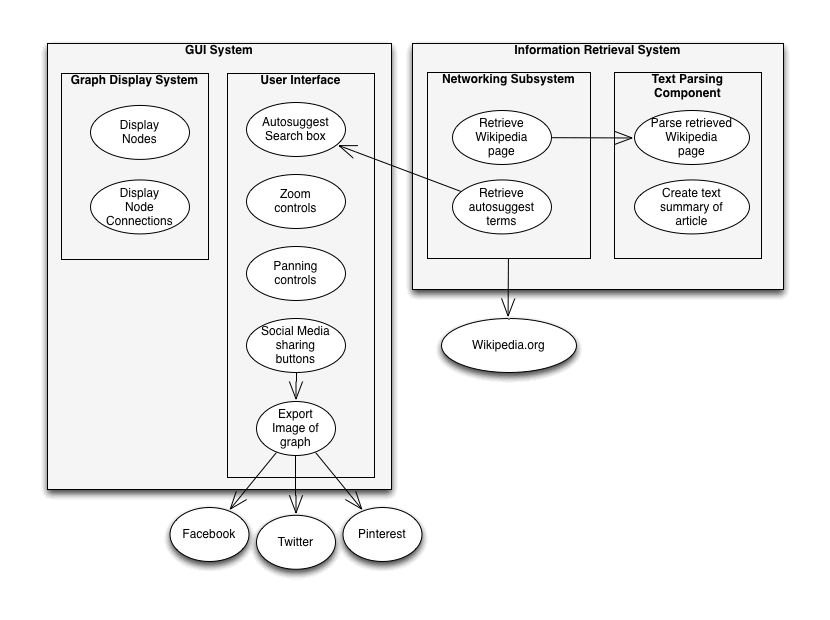
Our software will be in downloadable executable format. We will use Python for all of the application logic, and cross-platform libraries when possible to increase the ease of porting to other platforms in the future.

Minimum System Requirements:

OS: Windows 7

CPU: 2 Cores @ ~2GHZ

RAM: 2GB



## Design and Implementation Constraints

*Languages:* Python (networking, parsing, visualization)

*Libraries Used:*

GUI: kivy

Text Parsing: BeautifulSoup

Networking: urllib

*Hardware Limitations*: Network connection speed (accessing Wikipedia), memory (preloading related articles), processor and graphics card (visualizer)

*Other Factors:* Internet and en.wikipedia.org accessibility

## User Documentation

The Wikipedia Mapper UI is intentionally designed to be intuitive, and will therefore require little documentation. However, basic operational functions will be provided in a README file.

## Assumptions and Dependencies

*Assumption 1*: Reliable access to the internet in general and en.wikipedia.org specifically

*Assumption 2:* No major changes in the HTML structure of article pages

*Assumption 3:* Search term matches existing article title (partial or exact match)

*Assumption 4:* Related articles exist for a given search term

# **Specific Requirements**

## External Interface Requirements

### User Interfaces

The main elements of the GUI will be a search bar, graph display, and node manipulation. The search bar will consist of a text area for entering search terms and buttons to reset or initiate the search process. The graph display will contain an active article, including a text summary and a representative image from the Wikipedia page, if available. From this central point, the names of related articles will be displayed as link branches. Clicking on the name of a related article will spawn the branching nodes of that article; double-clicking a node reveals the text summary overlay. The user can also manipulate the node positions by clicking and dragging.

### Hardware Interfaces

None.

### Software Interfaces

Wikipedia Mapper will be built for Windows systems, though compatibility with Macs is also of interest. Several different kinds of Python libraries will be used by the software, including text summarization, graph visualization, manipulation and storage, GUI, and networking.

### Communications Interfaces

Wikipedia Mapper relies on a network connection with Wikipedia. In order to avoid over-taxing Wikipedia servers, throttle processes will be implemented to control the rate of connection requests. These processes may be customized for the network connection speed of individual systems (faster systems would be allowed more requests).

## Functional Requirements

* Enter search query

User will type their search term into a text area

* Clear search

Clear display area and reset search request

* Send queries to Wikipedia upon button click

Send contents of search text area to Wikipedia

* Throttle network requests

Control number of requests sent according to network connection speed

* Parse images, text, and links from corresponding Wikipedia article

Categorize article elements as image, text, or hyperlink

Extract groups of elements

Determine relationships between active article and related articles based on

hyperlinks

* Record query results as part of graph

Save elements in model

* Summarize article text

Use Python library to extract introductory paragraphs

* Draw graph structure in display area

Draw active article node with text summary

Draw branching related article nodes with their titles

* Show text summary and representative image upon double-clicking a node
* Change active article node upon clicking a related article node

Shrink active article node to title

Pan screen to center on relevant related article node

Render new related article nodes

* Reposition nodes

Alter position of nodes in display area upon clicking and dragging them

* Preload media elements

Save as many parsed elements and relationships to cache as possible for later

retrieval

* Future functionality

Import/export graphs

Connect and post to social media sites

Use Google as a data source

Apply same data parsing strategies to Google search queries

Switch between Google and Wikipedia search

## Behavior Requirements

### *Search* Use Case

1. User enters term in search box
2. User clicks “Search” button
3. Network connects to Wikipedia and retrieves the relevant article page
4. Parser finds related articles based on hyperlinks within page
5. Parser prioritizes related articles by relevancy
6. Model and UI create first active article node with number of related article nodes as specified by user
7. System stores current graph structure in model

**3.3.2 *Reset Search* Use Case**

1. User clicks on “Reset” button
2. Model deletes current graph structure
3. UI clears display area

**3.3.3 *Change Active Node* Use Case**

1. User clicks on a related article node
2. UI centers node on screen and sets corresponding article as the active article node
3. System performs Steps 5-8 under *Search* Use Case

**3.3.4 *Show Text Summary* Use Case**

1. Users double-clicks on the active node
2. Parser extracts introductory paragraphs from Wikipedia article
3. UI displays text summary overlay and representative image

//change for how you actually zoom

**3.3.9 *Zoom* Use Case**

# **Other Non-functional Requirements**

## Performance Requirements

Connections to Wikipedia and graph render time will not exceed 10 seconds

* Need to ensure network connection does not time out, and that the system does not try to render the graph for an indefinite period of time

## Safety and Security Requirements

None; no data is being sent across the network.

## Software Quality Attributes

* Interoperability: Software will have efficient two-way communicate with Wikipedia server .
* Security: Social network access will protect user log-in information.
* Usability: Users will be able to easily manipulate nodes and visually explore Wikipedia.
* Scalability: Software will perform well regardless of graph size and density.
* Performance: Software will access Wikipedia server and visualize articles quickly.
* (Optional) Portability: Software will have a Mac-compatible version.