Lab 1 - AskMissy Product Description

Matthew Stephens

Old Dominion University

CS 411W: Professional Workforce Development I

Professor Janet Brunelle

March 23, 2022

Version 2
Table of Contents

1. Introduction 3
2. Product Description 5
   2.1 Key Product Features and Capabilities 5
   2.2 Major Components (Hardware / Software) 8
3. Identification of Case Study 9
4. AskMissy Prototype Description 10
   4.1 Prototype Architecture (Hardware / Software) 11
   4.2 Prototype Features and Capabilities 13
   4.3 Prototype Development Challenges 15
5. Glossary 16
6. References 23

List of Figures

Figure 1: Major Functional Component Diagram 9
Figure 2: Prototype Major Functional Component Diagram 12

List of Tables

Table 1: Prototype Features Table 13
1. Introduction

Students face two large problems in the domain of reading. The first problem is that they lack easy access to relevant resources in relation to assignments they receive in school, notably in grades 9-12. The second problem is a general downward trend in the interest in reading for students of all ages. Whether the students come from historically underprivileged backgrounds, or they are exploring concepts that are underrepresented in traditional literature, a weak beginning in literature leads to a lack of interest, trust, and reliance on reading skills. Skills that many professionals find necessary for their day to day lives after their school careers (McLean).

In a 2007 study, researchers from the National Endowment of the Arts found that American students are reading less (Fuglei, 2019). From 1992 to 2002 students showed a 5% decline in extracurricular reading. This decline grew more severe for adults in this time frame, decreasing to 7%. Additionally, a study by the Pew Research Center found that adults who have lower levels of education-or are members of economically disadvantaged demographics-are less likely to read than adults with higher levels of education or economic advantage (Gelles-Watnick & Perrin, 2021). This shows a compounding effect where a low interest in reading at a younger age leads to further disinterest as an adult.

One way in which the harmful impact of a lower reading skill level is shown is in the Virginia Standards of Learning assessments, or SOLs (Sol Test Pass Rates & other results). These assessments are the state’s method of determining the learning and retention of the students in elementary, middle and high school for subjects such as math, reading, science and social studies. The curriculum of public schools is built around the SOL, and many students are taught that the SOL exam is the most important part of the school year as the culmination of all the lessons they have learned in that class. The Department of Education uses SOL pass rates,
among other statistics, to determine a school’s accreditation. When students do poorly on SOLs, schools are at risk of losing accreditation. This can cost schools valuable resources necessary to maintain or improve the student’s ability to learn.

When the COVID-19 pandemic slowed the in-person learning for schools in the 2019-2020 school year, students shifted to distance learning. They, and their parents, suddenly had to take on a much greater burden in terms of school organization and scheduling. A greater focus was put on independent study to keep up with the previous in-person learning standards. Simultaneously, teachers had to learn a completely new mode of teaching, disseminating lessons when they had difficulty ensuring that students were paying attention. Even when students returned to school for the 2020-2021 school year, the impact was clear by the drop in the SOL exam pass rate (Sol Test Pass Rates & other results). Math SOL pass rates dropped an average of 30% between the 2018-2019 and 2020-2021 school year. For reading SOLs in grades 3-12 the drop was lower at around 11%, though this is not cause for relief. Math skills can be buffered through memorization and practice of the same set of skills. Reading SOLs measure the test taker’s comprehension, which requires the intake of a wide range of books. Reading comprehension is another practice-based skill, but requires different learning strategies than math, which can be difficult when learning by oneself.

The drop in SOL pass rates exemplifies the need for all parties involved in primary education to restructure the learning process for students whose performance is non-optimized because of a lack of materials tailored to their learning preferences. Without this change, students lack exposure to the critical reading comprehension skills that will benefit them after their scholastic careers.
To address both the symptoms and core problems, AskMissy is a software application designed for students, teachers and librarians in grades six through twelve to help users find the resources necessary to bridge the gap between conventional and digital learning for a more personalized experience. Machine learning algorithms will be implemented to personalize the experience for users and predict the types of resources that the user would most benefit from.

2. Product Description

AskMissy is a web-based software application designed to help users in grades 9-12 find resources such as books to both aid them in their school careers and to invigorate their interest in reading. AskMissy will be more efficient in its recommendations by implementing machine learning algorithms to learn from the search results of the users over time and incorporate the learned preferences into future results. AskMissy databases will connect directly to the library databases of school libraries to allow users to plan out their resource acquisition through a number of options. Students will be able to search for books in the library based on several criteria such as book title, author, isbn number, or subject matter. Teachers will be able to upload the subject matter of their SOL curriculum and receive recommendations on resources that correspond. They will then be able to directly recommend those resources to their students who are organized in groups based on their classes. Librarians will maintain their libraries inventory and receive feedback from students and teachers for specific resources, as well as see statistics to reflect the popularity of those resources.

2.1 Key Product Features and Capabilities

AskMissy will be a web-based application that connects to the existing systems of each school that participates. The software would be centered around each specific school that has
registered to use it. Schools will be able to authenticate users either manually by way of an administrator or automatically connecting the registered user’s information to the school database of students, teachers or librarians. Once registered a user may access AskMissy anywhere so long as they have a device with internet access.

AskMissy will have two types of databases: the AskMissy Database and the School Library Database. All books stored in individual School Libraries will have their data drawn from the AskMissy Database, functioning as a central repository. The AskMissy Database is taken from the Goodbooks-10k GitHub repository. Unregistered users shall only have access to the AskMissy Database, while registered users shall view the books in their School Library first, as a means of verifying what books are most readily available.

Machine learning algorithms shall be implemented to both fill in the specific data for each resource in a school library as well as find resources that match a user’s interest and expand their horizons. Once AskMissy is connected to a school library database, the algorithm will go through the current inventory and connect those items based on their title, author and isbn number to the external database. Language recognition parses through the external database’s resources and draw primarily from the synopsis and reviews to find out what the subject matter of the resource is. This populates the tags for that resource which is stored in the AskMissy database. When a user searches for a resource, they will be given the chance to look for specific subject matter by entering a list of words. The algorithm then searches for resources whose tags most closely match the search criteria and return that resource to the user.

Users will be able to supply feedback to AskMissy based both on their overall impression of the resource and how well a given tag fits that resource. The feedback will take the form of a numerical rating of 1-5 inclusively, as well as the chance to submit a short (maximum 200
characters) written feedback. Once this feedback is received, AskMissy’s search algorithm will take this into consideration when supplying similar searches in the future, learning what resources best match any given search criteria.

A user who never registers or wants to access the software before logging in have limited access to the AskMissy software; a basic search function of the database with limited ability to input search criteria. Users who register have full access to the user-facing features of AskMissy, namely the machine learning search algorithms and group interactions. There are three different user types, but a single user may only ever be registered as a single type. The authentication process ensures that a user is not placed in the wrong category, but if an error does occur then the administrator will have access to correct this.

A student user may create a profile for themselves and input information about themselves to a limited extent. This limit is based around their school career, such as their grade or school club participation. Students will have access to the AskMissy search algorithm and can view their previous search results as well as the specific searches themselves. The specific amount of data saved for each student to view, and retrieve will be decided at a later date. They will be entered as a member of their specific class group as established by their teacher. They will have the option to view recommendations from other student users in their class and share their own recommendations. There will be limited ability to share details of their feedback that take the form of short (maximum 100 characters) reviews which other students may observe. Students may submit longer forms of feedback that are sent directly to higher tier users they are connected to, namely their teacher or one of the librarians in their school group. This takes the form of resource requests for items like books that are missing from their school library.
Teachers also have access to the AskMissy search algorithm. They are able to create class groups and connect their students into the group. Once a teacher has a group that teacher can then forward resources, they have found to their students either individually or in sets of students. These recommendations don’t affect the AskMissy search algorithm of any students until they submit feedback on that resource. Teachers may also submit resource requests to librarians. Teachers will have the option to create lesson plans in AskMissy based on the subject matter of each program. Teachers can then use AskMissy to provide them a list of resources that correspond to each lesson. Teachers will also be able to review the feedback that their students submit and run personalized reports on what students are searching and responding to.

Librarians will function similarly to teachers, though their groups will be school groups, and they will be connecting all registered teachers to the same group. They will also maintain the inventory of their school library. Librarians may receive the resource requests of the students and teachers. They may also directly contact administrators for bug support.

Administrators have access to all the features that the other user types have, as well as having access to the AskMissy data servers to ensure data integrity. They have access to all user profiles and user metadata. They are responsible for authorizing librarian users and connecting the AskMissy program to the school’s systems.

2.2 Major Components (Hardware / Software)

AskMissy requires the user to have access to some device capable of accessing the internet. On the backend, there are one or more servers dedicated to each of the following: The frontend website, the main backend database (holding user information and school library data), the web scraping process, the machine learning algorithm, and the external database (up to date data from web scraping). The full view of the components necessary is listed in Figure 1.
AskMissy is centered around the Python programming language and the application server will be decided at a later date, though most likely be chosen from Django, Gunicorn or Python Paste. The data that AskMissy collects from web scraping is stored in Flask file format.

The frontend of the website is written in a combination of HTML5, CSS3 and Javascript. The backend is python. The team’s programming IDE is PyCharm.

3. **Identification of Case Study**

AskMissy’s main intended users are students, teachers and librarians in grades 9-12 whose school curriculum is centered around the Virginia SOL. The schools in which AskMissy
is connected are some of the primary stakeholders, along with the parents of the students as secondary stakeholders.

The main goal of AskMissy is to supply a user with a tool that allows that user to obtain a list of resources tailored to their intended subject matter and interests. This goes beyond a customary search engine tool to look for a book on that user’s interests; instead, AskMissy learns how the user reacts to certain subjects and the quality of resources, then tailor future searches to focus on one type of resource, exclude others that offend their sensibilities, or open the user to a new world of subject matter in a way that encourages expanding horizons.

Other parties that benefit from AskMissy are publishers of resources who get overlooked in favor of more popular businesses. Because AskMissy is made to connect users to lesser-known resources, it follows that lesser-known publishers would glean a net positive from this as well.

4. AskMissy Prototype Description

The prototype contains most of AskMissy’s features, either fully or partially implemented. External sources like Amazon RDS are not implemented in the prototype due to a lack of knowledge of a workable execution. Security encryption is not implemented in the prototype, but login credentials are. All user roles have full functionality and testability in both the real-world product and the prototype, though the prototype will have a sixth user type in the Tester. The Tester is responsible for verifying the accuracy of the algorithm and the accessibility of the UI.
4.1 Prototype Architecture (Hardware / Software)

The web interface is available on desktop and laptop devices with internet connection. For users who choose to use their phones to participate in AskMissy, functionality is prioritized over the visual aspect. The prototype and its database are hosted on a physical server provided by ODU’s Computer Science department in a containerized environment using Docker. This environment also contains the MySQL server. There would be more separation between these resources in the real-world product to prevent a single failure point. Figure 2 shows a visual representation about the prototype implementation.
The AskMissy machine learning algorithm is written in the Python language using the PyCharm Integrated Development Environment, or IDE. Our database will be a Flask Server. Testing is done using unittest. GitHub will be utilized to keep track of which version of software is in use. User accounts are simulated to properly test the prototype and will be stored in the database.
4.2 Prototype Features and Capabilities

The prototype for AskMissy keeps most of the functionality as in the real-world product, but with a reduced scale. The application utilizes the AskMissy Search to personalize the recommended book to the individual user, so long as the user is logged in and is not a guest. The tag creation algorithm is implemented in the prototype to fill the test data set with workable tags for the AskMissy Search to use but is not implemented in the RWP. Communication between the user types is modeled to showcase the types of communication in the RWP. Users are organized into different classes and schools, with separated library catalogs to simulate access to different data sets. In the prototype the simulated users have simple profiles, only having enough details to test the AskMissy Search. The prototype will not have real connection to a school’s database of students, teachers, and librarians- instead, it provides a simulation of administrators authenticating the registration of the school side users. Users are able to view their previous searches and statistics that AskMissy gathers about themselves.

Table 1

AskMissy Real World Product vs. Prototype deliverables

<table>
<thead>
<tr>
<th>Category</th>
<th>Feature</th>
<th>RWP</th>
<th>Prototype</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Retrieval</td>
<td>Metadata Report</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
<tr>
<td></td>
<td>Basic Search</td>
<td>Full</td>
<td>Full</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AskMissy Search</td>
<td>Full</td>
<td>Full</td>
<td></td>
</tr>
<tr>
<td>Data Management - Live Product</td>
<td>Machine Learning</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>------</td>
<td>---------</td>
<td>------------------</td>
</tr>
<tr>
<td>Source Tag Creation</td>
<td>Full</td>
<td>Full</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source Tag Management</td>
<td>Full</td>
<td>Full</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesson Plans</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
<td></td>
</tr>
<tr>
<td>Internal Database Manipulation</td>
<td>Full</td>
<td>Full</td>
<td></td>
<td></td>
</tr>
<tr>
<td>External Database Manipulation</td>
<td>Full</td>
<td>Full</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Management - Development</td>
<td>Source Tag Development</td>
<td>None</td>
<td>Full</td>
<td>Use to develop default tags</td>
</tr>
<tr>
<td>Machine Learning Training</td>
<td>None</td>
<td>Full</td>
<td>Use to develop algorithm defaults</td>
<td></td>
</tr>
<tr>
<td>Simulated Data</td>
<td>None</td>
<td>Full</td>
<td>Use to fill database for testing</td>
<td></td>
</tr>
<tr>
<td>User testing reports</td>
<td>None</td>
<td>Full</td>
<td>Use to develop user interface</td>
<td></td>
</tr>
<tr>
<td>Security</td>
<td>Login/Authentication</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
<tr>
<td>Data Encryption, moving</td>
<td>Full</td>
<td>None</td>
<td>Best practices will be used</td>
<td></td>
</tr>
<tr>
<td>Data Encryption, resting</td>
<td>Full</td>
<td>None</td>
<td>Best practices will be used</td>
<td></td>
</tr>
<tr>
<td>Account Management</td>
<td>User Profile</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
<tr>
<td></td>
<td>Feedback</td>
<td>Full</td>
<td>Full</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group Management</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
<tr>
<td></td>
<td>Login/registration</td>
<td>Full</td>
<td>Full</td>
<td></td>
</tr>
<tr>
<td>UI</td>
<td>Group Interaction</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
<tr>
<td></td>
<td>Bug Report</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
<tr>
<td></td>
<td>Basic Search</td>
<td>Full</td>
<td>Full</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AskMissy Search</td>
<td>Full</td>
<td>Full</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
<tr>
<td></td>
<td>Personal Data Report</td>
<td>Full</td>
<td>Partial</td>
<td>Limited test data</td>
</tr>
</tbody>
</table>
4.3 Prototype Development Challenges

Prototyping the application involves several challenges. While most are planned for with mitigation techniques, some require handling as they arise. The AskMissy development team has little prior experience with Machine Learning algorithms and as such much of the functionality will need more rigorous testing and secondary input from the testing team to ensure quality. To mitigate this, the data being used will be as simple as possible while covering as many different scenarios as possible that would arise in the RWP. Similarly, connecting all the different databases presents as the largest hurdle. Again, sufficient testing will mitigate many of the issues that arise. Staggering feature development so that the backend is complete before the frontend provides more opportunity for the frontend development to go smoothly, connecting to an existing database without having to change the implementation functions more than needed.
5. **Glossary**

**Administrator**: A user who is responsible for managing a majority of AskMissy’s working data.

**Agile**: A set of frameworks and practices where solutions evolve through collaboration between self-organizing cross-functional teams.

**Acknowledgement**: A message delivered to an authenticated user in response to their submitted bug report.

**Activity**: Any action undertaken by a user in relation to the AskMissy application.

**Announcement**: A message delivered to lower-level users from a higher level user, usually in the case of a Librarian or Teacher to a student.

**AskMissy**: A software application that will help users find more relevant resources.

**AskMissy Library**: The total collection of metadata from which other libraries and functions Extract data from.

**Application Programming Interface (API)**: A software intermediary that allows two applications to talk to each other.

**Apriori Algorithm**: An algorithm designed to find Itemsets in a dataset for Boolean association rules. Itemsets are groups of books found to be read together with a high frequency, implying similar future association.

**Authenticated User**: A user who possesses an account in AskMissy, i.e. not a guest.

**Association Rule (Algorithm)**: A statement that a book or group of books implies the presence of another item with some probability.
Basic Search: A search function that lists the highest rated books based on the search criteria, including genre, author, title, and publication date. This search does not utilize the Apriori Algorithm or any specific school library.

Book Data: Data about a book entry in either a School Library or the AskMissy Library, describing the book’s title, author, isbn number, genres, average rating, number of ratings, publication date, original title (if any), and language.

Bug: An error in the AskMissy application that causes it to produce an incorrect result or behave in unintended ways.

Categorize: A feature of the bug reporting system which lets a user assign greater specificity to the nature of a bug.

Classes: The sections of a course that are scheduled for a specific academic year, assigned to Teachers, and include a roster of Students.

Confidence (Algorithm): The ratio of transactions that contain book A and B to transactions that contain book A.

Conviction (Algorithm): The ratio of expected support of book A occurring without book B assuming that books A and B are independent, to the observed support of A occurring without Y.

Courses: The programs of study which the Authenticated Users of AskMissy are enrolled in.

Cascading Style Sheet Revision 5 (CSS5): A style sheet language used for the presentation of documents written in a markup language such as HTML, CSS5 is the fifth version of the original CSS version.

Comma-Separated Values (.CSV): A delimited text file that utilizes commas to separate values.
Current Books: The list of books that an Authenticated User has declared they are actively reading.

Data Retention: The continued storage of an organization's data for compliance or business reasons.

Database: An organized collection of structured information, data, typically stored in a computer system.

Economically Disadvantaged: A student eligible for Free/Reduced Meals who receives Temporary Assistance for Needy Families (TANF) or is eligible for Medicaid.

Exact Match Search: A search for a single specific type of resource.

Extract: To receive or collect data from a data source, usually one of the library databases associated with AskMissy.

Federal School Code: A six-digit character code to identify a specific school or educational institute.

Flask: A micro web framework primarily written in Python.

File Server: A device that controls access to separately stored files.

Filter: To specify the results to view from an inquiry.

Goodreads: A subsidiary database of Amazon that stores books, annotations, quotes, and reviews.

Group: A collection of users organized into two possible levels - Classroom or School. Students and Teachers will be part of a classroom group and a school group, Librarians will be part of a school group.

Guest: A user who is not a Student, Teacher, Librarian, or Administrator, who has limited access to the AskMissy application.
**Hypertext Markup Language Revision 5 (HTML5):** A type of markup language primarily used for implementing content in the World Wide Web, this is the fifth version of its original version.

**Input:** Supplying data to the AskMissy application, or the data being supplied.

**Interaction:** The means by which one user may share information or otherwise communicate with another user. This may be done predominantly through the use of Messages, unless otherwise specified.

**Interests:** Aspect of the User Profile based on the user’s liked books.

**Itemsets (Algorithm):** a grouping of books found to be associated with each other across multiple user’s past reading.

**Lesson Plan:** Input supplied by Teachers to the AskMissy application to provide a template set of search parameters which other Authenticated Users may use to perform an AskMissy Search, usually relating to one or more Courses.

**Librarian:** A user responsible for managing the library’s inventory/database, communicating with teachers and students.

**Lift (Algorithm):** The ratio of observed support of book A and B to the expected support of book A and B.

**List (Algorithm):** a comma delimited file (.csv) consisting of one or more columns containing one or more entries in the format of rows, with each data form separated by a comma.

**JavaScript:** A programming language that is used for implementing websites on the World Wide Web.

**Message:** A communication in the form of a string data type between one or more users.

**Metadata:** Data that provides information about other data.
**Personal Learning:** An educational approach that aims to customize learning for each user’s strengths, needs, skills, and interests.

**Profile:** The displayed data for an authenticated user. This data describes the user’s type,

**Private (Data):** Authenticated User Activity which cannot be viewed by other Authenticated Users.

**Public (Data):** Authenticated User Activity which can be viewed by other Authenticated Users.

**Python:** A high-level programming language.

**Query:** An action functions perform to obtain data corresponding to one or more Activities the user is performing.

**Rate:** A numeric measure of the quality of any given book on a scale from 1 (lowest quality) to 5 (highest quality).

**Recommendation:** A specific book that a Teacher or Librarian may submit for Students to view.

**Request:** A Message from a Student to a Teacher or Librarian specifically to bring attention to the Student’s desire for a book to be included in the School Library. Individual Requests can also be sent from the Teacher to the Librarian to emphasize the importance of that Request.

**Real World Product (RWP):** Refers to the physical version of any digital abstraction described within the AskMissy documentation.

**Register:** The process that a guest user takes to create an account verified by their school’s database.

**Review (Communication):** A text assessment submitted by an Authenticated User regarding a particular book.
Review (Administrator): An activity the Administrator User can perform to access the AskMissy data in any form and make minor modifications according to the context of the documentation.

School Library: The database of books registered in the Real-World school.

Standards of Learning (SOL): An examination conducted by Virginia Public Schools that tests the minimum required expectations for every student enrolled in the state of Virginia.

Student: A user studying at a K-12 education institution.

Student Feed: A portion of the user interface that allows student users to view the most recent books read and/or reviewed by their fellow students, and view books specifically recommended by the Teacher assigned to their class.

Support (Algorithm): The ratio of transactions that contain an itemset to all transactions.

Shelves: Term used in the Test Library for attributes that describe a book’s subject matter and metadata. Synonymous with Tags.

Short Message: A Message specifically no longer than 200 characters, spaces included.

Submit: The process of Inputting a required data type for the intended process.

Tags: Term used for attributes that describe a book’s subject matter and metadata. Synonymous with Shelves.

Teacher: A user who helps K-12 students acquire knowledge. They are responsible for making plans and managing students’ groups/communication.

Teacher Feed: A portion of the user interface that allows teachers to view the most recent books read by that teacher’s students, and view recommendations by other Teachers and Librarians.
Temporary Assistance for Needy Families (TANF): A program that provides eligible families with a monthly cash payment to meet their basic needs.

Test Library: Database of books, users, shelves/tags, and ratings drawn from the goodbooks-10k GitHub repository.

Tester: A user responsible for designing and conducting testing suites for usability testing.

User: An individual using the AskMissy Interface.

User Interface/User Experience (UI/UX): The visual representation of the data AskMissy provides to the user on the user’s computer.

View: The current information being displayed in the UI/UX to the user, or the Activity of interacting with the UI/UX.

Web Scraping: The process of extracting content and data from a website.

Web Server: A computer program that distributes web pages as they are requisitioned.

Windows: A series of operating systems developed by Microsoft.
6. References


