Lab 1 – AskMissy Product Description

Tanner Henderson

Old Dominion University

CS 411W

Professor Janet Brunelle

March 24, 2022

Version 2

Table of Contents

List of Figures

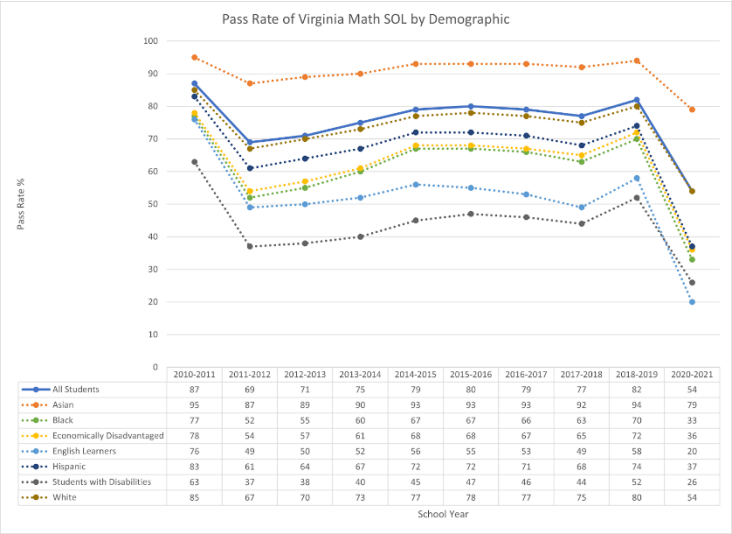
List of Tables

1 Introduction

Within society, attempting to persuade students to read is a difficult problem to undertake. This problem has only been getting progressively more difficult as students have recently been transitioning into the classroom through virtual means with the onset of the recent COVID-19 pandemic. This lack of reading presents a large problem in modern society, as reading influences subjects and students' grades or performance with them. The problem of lack of reading due to difficulty finding resources is apparent through SOL, or the Virginia Standards of Learning, scores that have been gathered. These SOL scores, seen in Figure 1, show that students have been largely underperforming due to the effects that COVID-19 has had on the population and the implementation of online learning.

Figure 1

Math SOL by Year. (Virginia Department of Education, 2021)



	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019	2020-2021
All Students	87	69	71	75	79	80	79	77	82	54
Asian	95	87	89	90	93	93	93	92	94	79
Black	77	52	55	60	67	67	66	63	70	33
Economically Disadvantaged	78	54	57	61	68	68	67	65	72	36
English Learners	76	49	50	52	56	55	53	49	58	20
Hispanic	83	61	64	67	72	72	71	68	74	37
Students with Disabilities	63	37	38	40	45	47	46	44	52	26
White	85	67	70	73	77	78	77	75	80	54

The lack of resources available through searching, either by teacher or student, have impacted interest in learning, as well as the lowering of SOL scores, and is a problem that needs to be fixed in society when wishing to better help school systems. COVID-19 affects reading and thus affects the SOL scores of students and their given schools, and a solution to the problem,

that problem being that students' performances are non-optimized due to lack of materials and resources available to students through online learning, is needed. This solution is what the AskMissy software will provide to the educational system.

The AskMissy will help its users, primarily being users within a school system, to find materials and resources, such as books and supplemental instructional materials, necessary to bridge the gap between conventional and virtual learning for a more personalized experience for each user. The AskMissy solution is a web application that finds resources for its users, specifically personalized based on their own preferences, as well as being based upon feedback they give on a given resource. AskMissy is intended for middle to high school students, or grades 6-12, as those students would best benefit from this software. AskMissy shall implement a machine learning algorithm to allow better recommendations on what a given user may be preferential toward based on their user profiles and their previous feedback on resources that AskMissy recommended. The software will also allow teachers and librarians to input their lesson plans into the software in order to have a resource recommended to students for each lesson plan.

2 Product Description

AskMissy is a software application that has the purpose of allowing more preferential recommending of resources for users and allowing teachers and librarians to have resources entailed to specific lesson plans for students. It is a web-based application and utilizes a machine learning algorithm to constantly improve future recommendations for users, as well as being an ever-learning software that adapts based on each specific user. The primary goal in creating AskMissy is to improve resource finding capabilities for students, teachers, and librarians. A

secondary goal to be achieved, is that this software promotes higher SOL preparation in the future for students. There is also a goal that the SOL preparation eventually extends to the teachers and librarians, as they are also a crucial aspect of student SOL preparation and recommending students suitable resources is a priority for the software.

2.1 Key Product Features and Capabilities

The initial idea of AskMissy is that the main functions of the software will be located in a school system's network and will be accessible via web application to users that have valid school identification numbers. However, AskMissy shall be available for access by authenticated users and guests both inside and outside the school's network.

Authentication and user differentiation is a large aspect of the AskMissy software. Non-registered users, or guests, within the AskMissy system will be allowed to access the system, though only to a limited extent. This limited extent shall only be a basic search functionality, similar to searching a standard library database, and the ability to create a registered profile within the system. Having limited functionality for guests will reduce the risk of a potential malicious user ruining ratings or algorithm searches/tags within AskMissy, if they are not the intended audience of the program. Registered users, on the other hand, will have access to the AskMissy system and its search and feedback algorithms. The authentication of these users shall be through the use of school identification numbers, to verify they are a valid school user, and shall be verified with general ease since the application will be placed and positioned within a school system. These registered users are students, teachers, librarians, and administrators, each with differing degrees of permissions and capabilities.

Students will have access to the AskMissy search algorithm in order to find resources personalized to them. They shall also have the ability to share resources with other users on the platform, such as sharing a book that was recommended to them to another student or a teacher. They will also have the ability to message higher-tier users, such as teachers. The message functionality shall be used to ask for a resource that is not available in the library or to ask teachers or librarians for specific resources that they may be preferential toward. The students shall have limited access within the group or courses that they are apart of, such as joining, leaving, or accessing reviews from the teachers who are the owners of the group/course.

Teachers will be a higher-tier user than students and will thus have all the permissions that a student would have with specific additions. The additional permissions include receiving resource requests from students, sending resource requests to librarians, and course group creation and management. The management of course groups will also provide them the ability to remove students from their course group and will allow them to have access to students' reviews of resources that are made available within their course group. Teachers will also have the ability to create a lesson plan with specific resources attached to them, to allow for easy shareability with other teachers or students.

Librarians will have access to all student and teachers features, except the teacher ability to make and manage a course group. The librarian does not have a course group, so they will instead be able to create and manage a school group in a similar manner, which contains students, teachers, and courses. Librarians will also have the ability to maintain and update the school library inventory within the AskMissy program to allow the program to remain as current as possible. They can also manage lists of requests from teachers with a priority order list for

resources that have been requested by students or teachers. They are also able to contact administrators for technical support on the system.

Administrators in AskMissy will have all accesses of the previous user types. The main role and features of the administrators are to ensure data integrity and maintain databases within the program and will thus be given access to all user profiles and metadata and will be allowed limited database management capabilities. They will also be able to authorize librarians to create school groups within the system and contact the administrators of the school for data requests.

Feedback is another major function and aspect of the AskMissy solution. Feedback of a resource will be requested by the software after that resource was recommended to the user. The software will prompt the user, asking for their feedback; this feedback will then be used in the machine learning algorithm to optimize the results in the future. An important note is that within the AskMissy software, overall ratings of a resource will not actually affect a user's personal profile recommendations, thus keeping the initial idea of personalized searching intact.

The databases within AskMissy are the primary concern aside from machine learning. The school database, comprised of the school library and other resources, will be available with to AskMissy through its integration within the school system, and these databases will be edited and maintained by the librarian and administrator user types. External databases will also be a feature of AskMissy and will be accessed by web scraping algorithms and APIs to allow for more resources to recommend to prospecting users. There is also the AskMissy user profile database which will store all users, their preferences, and their tags.

Another main aspect of AskMissy is the machine learning algorithm that will be implemented. The machine learning algorithm will allow for personalized search results found

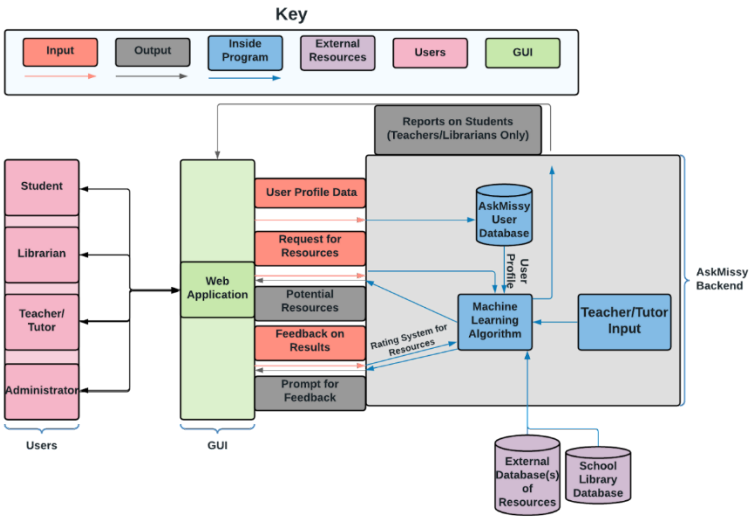
for each specific user based on the data acquired from account creation and feedback received about past recommended resources from AskMissy. This machine learning algorithm will allow the product to constantly grow alongside its users for a more optimized experience. It is intended for AskMissy to be more optimized each time it is used, meaning more use of the software leads to more preferential results for its users.

2.2 Major Components (Hardware/Software)

AskMissy has both front-end user interfaces and backend servers, databases, and components. Figure 2 shows the major functional component diagram, or MFCD, for AskMissy, which will detail the software and what features are within the AskMissy application.

Figure 2

AskMissy Major Functional Component Diagram



AskMissy needs hardware in order to run effectively. The hardware to access the AskMissy application, however, only needs a device capable of accessing the internet, as AskMissy will be a web application. The servers needed to properly run the AskMissy program

and application is a main frontend server, a main backend server, a web scraping backend server, a machine learning backend server, and a server dedicated for the main database.

For software, the frontend software will be programmed in the languages of HTML5, CSS3, and JavaScript, to allow developers to interface as a web application effectively. The backend will primarily be programmed using the Python language, due to Python's extensive machine learning based libraries available for use. The development environment that will be primarily used for the AskMissy development will be PyCharm. Automatic navigation libraries will be done through the use of PyAutoGUI. The repository that the development team will use to version control the current build of the project will be GitLab, allowing updates and forks to the project when necessary. The databases within AskMissy will be using standard choices such as Amazon RDS, MySQL, and connector/python. The machine learning, which is a large part of the backend, will be programmed using Python, specifically using the library of scikit-learn, to properly utilize machine learning based functions. The natural language processing algorithm will also be programmed using Python.

3 Identification of Case Study

The AskMissy product is intended for students, teachers, and librarians, and is primarily intended for middle and high school students between grades 6-12. The product initially intends to increase SOL scores for students in the future but can be expanded on its intended use as well. AskMissy is intended, in general, for assisting students in searching for personalized resources to make their experience with gathering resources more optimized and particularly making said resources pertain to their interests, allowing students to become more invested. The AskMissy

program will also be used for teachers and librarians in order to enable them to find appropriate resources for themselves, their lessons, and their students.

As AskMissy grows – in addition to assisting these students, teachers, and librarians with their academic careers – it will also have a potential to benefit resource publishers, as they would receive additional sales of their resources if requested resources are not available in the school library. The software will also help students' parents, as they will also benefit from their child's success and accomplishments.

4 AskMissy Product Prototype Description

The prototype for AskMissy will be a scaled-down yet functional version of the fully realized and developed AskMissy product, allowing for the main innovational aspects of the AskMissy product to be realized without having to fully develop parts of the software that will use standard practice techniques.

The prototype will still allow for each variant of user, those being guests, students, teachers, librarians, and administrators, and will have an additional user and role of tester, which will test the prototype to ensure its functionalities are intact and will have capabilities of all other users. The tester also has the ability to create or change data that would otherwise be immutable in the real-world product of AskMissy, in order to simulate testing criteria.

The AskMissy prototype will have the standard and personalized search function implemented as it is the primary function of AskMissy. The machine learning algorithm within the prototype will be a pseudo-machine learning due to time constraints of implementing a real machine learning algorithm. The pseudo-machine learning algorithm will be largely logic-based to allow it to be comparable to a standardized machine learning algorithm. The prototype will

have simulated accounts and data created and placed within its databases in order to fully test the application and its algorithms. The amount of data that will be available will be less than the real-world product but will be sufficient for the needs of the testers to fully develop and test the algorithms used within AskMissy.

The security within the prototype of AskMissy will largely not be implemented, as it is typical to use standard practice for securing data within the system and information is available on this, meaning that it is not relevant specifically to the product and would hamper already constrained development time.

4.1 Prototype Architecture (Hardware/Software)

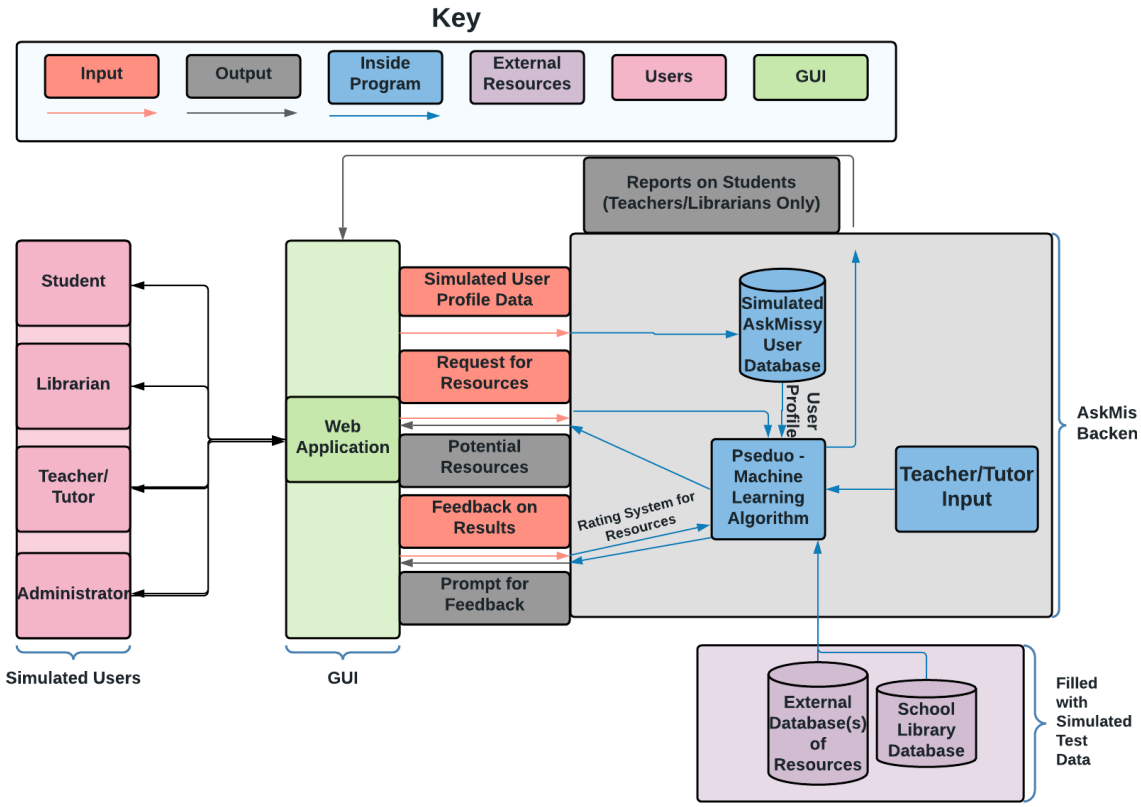
The AskMissy prototype will also have slightly differing hardware and software requirements. The hardware utilized within the AskMissy prototype will mainly be the ODU Computer Science Department's virtual machine. This hardware will allow hosting of servers for databases and the web application. Aside from the virtual machine, there will be no other specialized hardware in order to develop and run the AskMissy prototype, only needing a standard computer with internet access. However, the software required for the prototype is larger.

For version control of the prototype code and database scripts, Git will be used with a personalized Git repository for the development team of the AskMissy prototype. Development of said code and algorithms will be done using the IDE of PyCharm and using the Python programming language for its pseudo-machine learning algorithms due to its extensive libraries involving machine learning. The pseudo-machine learning algorithm will be using implementations of the libraries Pandas, NumPy, SciPy, Matplotlib, and Jupyter Notebook. The

front-end of the prototype will be implemented using HTML, CSS, and JavaScript to develop the webpage. The backend database structure shall use MySQL for the AskMissy prototype. Figure 3 shows is the MFCD for this prototype.

Figure 3

AskMissy Prototype Major Functional Component Diagram.



4.2 Prototype Features and Capabilities

The AskMissy prototype will retain all core functionalities of the real-world product of AskMissy, including personalized recommendations and the enabling of inter-user resource sharing. The utilization of a tag-creation algorithms will be implemented specifically in the prototype, but not implemented in the real-world product. The simulated data and user testing reports will also not remain when being developed as the real-world product and will only be an

available function in the prototype. A list of all changes between the AskMissy prototype vs. its real-world product counterpart is shown in Table 1.

Table 1

AskMissy Real-World Product vs. Prototype Table

Category	Feature	RWP	Prototype	Reasoning
Data Retrieval	Metadata Report	Full	Partial	Limited test data as a proof of concept
	Basic Search	Full	Full	
	AskMissy Search	Full	Full	
Data Management - Live Product	Machine Learning	Full	Partial	Limited test data as a proof of concept
	Source Tag Creation	Full	Full	
	Source Tag Management	Full	Full	
	Lesson Plans	Full	Partial	Limited test data as a proof of concept
	Internal Database Manipulation	Full	Full	
	External Database Manipulation	Full	Full	
Data Management - Development	Source Tag Development	None	Full	Use to develop default tags
	Machine Learning Training	None	Full	Use to develop algorithm defaults
	Simulated Data	None	Full	Use to fill database with simulated data for testing
	User testing reports	None	Full	Use to develop user interface
Security	Login/Authentication	Full	Partial	Limited test data as a proof of concept
	Data Encryption, moving	Full	None	Best practices will be put in place
	Data Encryption, resting	Full	None	Best practices will be put in place
Account Management	User Profile	Full	Partial	Limited test data as a proof of concept
	Feedback	Full	Full	
	Group Management	Full	Partial	Limited test data as a proof of concept
	Login/registration	Full	Full	
UI	Group Interaction	Full	Partial	Limited test data as a proof of concept
	Bug Report	Full	Partial	Limited test data as a proof of concept
	Basic Search	Full	Full	
	AskMissy Search	Full	Full	
	Communication	Full	Partial	Limited test data as a proof of concept
	Personal Data Report	Full	Partial	Limited test data as a proof of concept

4.3 Prototype Development Challenges

There are expected challenges the development team must overcome when developing the AskMissy prototype to be functional and ready to present in the designated time available. One large challenge for the team going forward is the lack of knowledge and experience in certain aspects and areas of the development process. Recently having lost the team's lead subject matter expert on machine learning and algorithms has made this challenge even more

prevalent for the algorithm development team. To couple with this issue, most of the team is also learning new development tools alongside this, meaning that learning sufficient skills to achieve a functional prototype in the time allotted is a difficult hurdle to overcome. There is also the challenge pertaining to the simulated data for the prototype. The leading concern in this area is that there is a large amount of data that must be simulated in order to fully test algorithms that are developed and multiple databases with a variety of tables to store said data. With a surplus of data that is needed, the problem of where this data comes from arises, as well as the problem of properly being able to parse through said data to obtain the information that is 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

Allreaders.com features detailed book and movie reviews from many different genres of books! Detailed Book review summaries. (n.d.). Retrieved December 14, 2021, from http://allreaders.com/

Amazon. (2016). *Amazon.com: Kindle eBooks: Kindle Store: Nonfiction, Literature & Fiction, Foreign Languages, Business & Money & More*. Amazon. Retrieved December 14, 2021, from https://www.amazon.com/Kindle-eBooks

Education, V. D. of. (n.d.). *Sol Test Pass Rates & other results*. VDOE :: SOL Pass Rates Results & Other Results. Retrieved December 14, 2021, from https://www.doe.virginia.gov/statistics_reports/sol-pass-rates/index.shtml

Education, V. D. of. (n.d.). *Sol Test Pass Rates & other results*. VDOE :: SOL Pass Rates Results & Other Results. Retrieved December 14, 2021, from https://www.doe.virginia.gov/statistics_reports/sol-pass-rates/index.shtml

Fuglei, M. (2019, July 22). *Why students who read for pleasure are stronger academically*. Resilient Educator. Retrieved December 14, 2021, from https://resilienteducator.com/classroom-resources/how-reading-for-pleasure-helps-students-develop-academically/

Garcia, E., Weiss, E., & Welshans, I. (2020, October 7). *What teaching is like during the pandemic-and a reminder that listening to teachers is critical to solving the challenges the coronavirus has brought to public education*. Economic Policy Institute. Retrieved December 14, 2021, from https://www.epi.org/blog/what-teaching-is-like-during-the-pandemic-and-a-reminder-

that-listening-to-teachers-is-critical-to-solving-the-challenges-the-coronavirus-has-brought-to-public-education/

Gelles-Watnick, R., & Perrin, A. (2021, November 1). *Who doesn't read books in America?* Pew Research Center. Retrieved December 14, 2021, from https://www.pewresearch.org/fact-tank/2021/09/21/who-doesnt-read-books-in-america

Goodreads. (n.d.). *Meet your next favorite book*. Goodreads. Retrieved December 14, 2021, from https://www.goodreads.com/

Ingraham, C. (2021, November 24). *Leisure reading in the U.S. is at an all-time low*. The Washington Post. Retrieved December 14, 2021, from https://www.washingtonpost.com/news/wonk/wp/2018/06/29/leisure-reading-in-the-u-s-is-at-an-all-time-low/

What should I read next? book recommendations from readers like you. (n.d.). Retrieved December 14, 2021, from https://www.whatshouldireadnext.com/

Rosalina, E. & Nasrullah. (2019). The correlation between self-esteem and student's reading comprehension. English Language Teaching Educational Journal, 2(2), 70-78.