



**UNIVERSITY OF THE PHILIPPINES
OPEN UNIVERSITY**

MASTER OF INFORMATION SYSTEMS

MIGUEL V. ABRIOL-SANTOS

**AUTOMATED CATEGORIZATION OF ARTICLES USING NATURAL LANGUAGE
PROCESSING**

Adviser:

KATRINA JOY M. ABRIOL-SANTOS
Faculty of Information and Communication Studies

01 JANUARY 2060

Permission of the classification of this academic work access is subject to the provisions of applicable laws, the provisions of the UP IPR policy and any contractual obligations:

Invention (I)	YES/NO
Publication (P)	YES/NO
Confidential (C)	YES/NO
Free (F)	YES/NO

Student's signature:

Adviser's signature:

University Permission Page

AUTOMATED CATEGORIZATION OF ARTICLES USING NATURAL LANGUAGE PROCESSING

I hereby grant the University of the Philippines a non-exclusive, worldwide, royalty-free license to reproduce, publish and publicly distribute copies of this Academic Work in whatever form subject to the provisions of applicable laws, the provisions of the UP IPR policy and any contractual obligations, as well as more specific permission marking on the Title Page.”

I specifically allow the University to:

1. Upload a copy of the work in the theses database of the college/school/institute/department and in any other databases available on the public internet
2. Publish the work in the college/school/institute/department journal, both in print and electronic or digital format and online; and
3. Give open access to the work, thus allowing “fair use” of the work in accordance with the provision of the Intellectual Property Code of the Philippines (Republic Act No. 8293), especially for teaching, scholarly and research purposes.

MIGUEL V. ABRIOL-SANTOS/ Date
Signature over student name and date

Approval Page

This paper prepared by MIGUEL V. ABRIOL-SANTOS with the title: "AUTOMATED CATEGORIZATION OF ARTICLES USING NATURAL LANGUAGE PROCESSING" is hereby accepted by the Faculty of Information and Communication Studies, U.P. Open University, in partial fulfillment of the requirements for the degree MASTER OF INFORMATION SYSTEMS.

KATRINA JOY M. ABRIOL-SANTOS

Adviser

Date Signed

DR. RIA MAE H. BORRAMEO

MIS Program Chair

Date Signed

DR. DIEGO S. MARANAN

Dean

Faculty of Information and Communication Studies

Date Signed

ACKNOWLEDGEMENT

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. MUST NOT EXCEED ONE PAGE.

ABSTRACT

Please use this as a guide for writing your abstract. Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. **MUST NOT EXCEED ONE PAGE.**

TABLE OF CONTENTS

	<u>PAGE</u>
TITLE PAGE	i
APPROVAL PAGE	ii
APPROVAL PAGE	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	vii
LIST OF FIGURES	viii
INTRODUCTION	1
Background of the project	1
Objectives	1
Scope and Limitations	2
Scope	2
Limitations	2
REVIEW OF EXISTING ALTERNATIVES	4
How to cite references	4
Apa Subsubsection	4
Why do we need this?	4
PROJECT DETAILS	7
Overview	7
Project Framework	8
Materials/Technologies Used	8
System Design	8
Implementation	8
Project Assessment	9
User Testing	9
Security Testing	9
RESULTS AND DISCUSSION	10
SUMMARY AND CONCLUSION	12

FUTURE WORK	13
LITERATURE CITED	14

LIST OF TABLES

<u>TABLE</u>		<u>PAGE</u>
1	Summary of Existing Alternatives	5
2	Test Caption Table	10

LIST OF FIGURES

<u>FIGURE</u>		<u>PAGE</u>
1	Test Caption Figure	7

INTRODUCTION

Background of the project

Your introduction must address the following questions (note that this is a guide, the questions must not be the subsections in the Introduction):

- What is the problem that you are trying to address?
- Why is the problem interesting?
- What is the solution of others? Include a summary of what the others are currently doing.
- What is the background of the proposed solution?
- What is being proposed in this document? Include the following: summary of methodology and the expected results

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Objectives

What do you wish to achieve in your capstone project? Write one general and at least three specific objectives of your project. The objectives must:

- be clear, concise, declarative statement;
- provide direction to investigate the variables in the study;

- focus on ways to measure the variables, to identify or describe them;
- identify relationships between variables;
- indicate results sought by the project proponent at the end of the process;
- be closely related to the statement of the problem; and
- be SMART: Specific, Measurable, Attainable, Realistic, Time-bound.

Scope and Limitations

In a research or capstone project, the scope and limitations are essential elements that guide the work to stay focused and define the extent of the content to be covered.

Scope

The scope of a study refers to the extent or boundary of the research and defines where the project will be deep-diving. It outlines what the project will encompass and also determines its size and the parameters to work within, including specific objectives, activities, resources, methodologies, and deliverables. For instance, if the project is about developing an AI model to predict stock prices, then the scope could include analyzing given datasets, pattern recognition, implementing machine learning algorithms, testing the model's accuracy, and more.

In terms of scope, a clear definition is a vital step to avoid research deviations. It ensures that the project remains on track without straying into unrelated areas. If the scope is not well defined, the study can easily grow beyond your control.

Limitations

The limitations of a study are potential weaknesses or problems encountered during the study due to constraints on methodology or resources. These might include a lack of available or reliable data, time or financial constraints, lack of access to technology or

software, and the assumptions you've had to make due to such constraints.

For example, in the above-mentioned AI model, a limitation could be the unavailability of real-time stock data for accurate testing, or perhaps certain algorithms couldn't be used due to software limitations.

REVIEW OF EXISTING ALTERNATIVES

This chapter should be able to present the following in not more than 600 words: description of how users/clientele currently cope with the problem; assessment of the best available resources for addressing the problem; and description of how your project stands out in comparison to the existing alternatives.

How to cite references

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Apa Subsubsection

APA CITATION SAMPLE HERE USING BIBTEX (Crescenzi & Kann, 1997)

Why do we need this?

The review of existing alternatives is a thorough study of existing applications or solutions and literature that are related to your project. We do this for the following reasons:

- you get to see what other professionals or researchers and authors in the area where your topic falls have done or are currently doing;
- you will have a deeper understanding of your topic and of the application that you are proposing to develop;
- you will know if someone else has already done the same (if you are duplicating a

project or a research); and

- you will have an idea on how other people solved the problems that are similar to yours.

Instructions: Fill in the table below with your proposed system and at least 3 related systems. Be concise and critical. After completing the table, answer the reflection questions. Once all your answers are in place, remove the instructions and examples before compiling and submitting.

Table 1. Summary of Existing Alternatives

System (License)	Platform	Brief Description	Basic Functionalities	Relation to the Proposed System
Proposed System: StudyBuddy AR (to be determined)	Mobile (Android/iOS)	AR app for visualizing academic concepts across disciplines	View molecules, anatomy, circuits, and diagrams in AR; interactive learning modules	–
Google Expeditions (Free but discontinued)	Mobile (Android/iOS)	AR/VR app for virtual field trips and basic concepts	Immersive AR/VR lessons and tours	Both use AR for education but Expeditions is broad and discontinued
Anatomy 3D Atlas (Paid - Subscription)	Mobile (Android/iOS)	Interactive 3D atlas for studying human anatomy	Layered 3D models of organs and systems	Similar in 3D visualization but limited to medicine
AR Circuits (Free)	Mobile (Android)	Educational AR app for visualizing circuit components	Place circuit components in AR and test connections	Similar AR approach but restricted to electronics

Discuss the details of each existing alternative and discuss this question: Can You Use This Instead? Why or Why Not? In addition, discuss why your proposed system is necessary, focusing on the following:

- Biggest gap in existing systems: Example: No existing AR learning app provides a single platform covering multiple subjects. Most are subject-specific.
- Why our project is still worth developing: Example: StudyBuddy AR combines features from various niche apps into one unified platform, making it more versatile for students.

- Our unique contribution: Example: A cross-disciplinary AR educational tool designed for general college learning, not restricted to one field.

PROJECT DETAILS

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio. Quisque ullamcorper placerat ipsum. Cras nibh.



Figure 1. Test Caption Figure

Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Overview

This is a brief description of the proposed system.

Project Framework

A project framework is the overarching structure or outline that guides the planning, execution, and management of the project.

Materials/Technologies Used

This includes the technologies to be used or used in the development of the system and their descriptions, and how they are utilized in the project.

System Design

Include functional and non-functional requirements and their descriptions, user case design, database design, etc.

Implementation

The details of the implementation of the system. For IS 295A, you need to include an implementation plan and include a Gantt chart as one of the appendices. For IS 295B, include screenshots of the system and their descriptions.

Project Assessment

User Testing

- How did you test the system?
- How did you find bugs?
- Give all the details: the test specifications, the number of users who tested it, etc.
- Explain results. How did you fix the bugs?
- Did you do regression testing after fixing the bugs?

Security Testing

Use this tool for testing the security of your system. What are the security vulnerabilities of your system? How did you/will you address that?

RESULTS AND DISCUSSION

What are the things you learned as you did this project? What are the challenges that you encountered? Share relevant insights. Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Table 2. Test Caption Table

1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus

aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi. Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

SUMMARY AND CONCLUSION

Here's a guide on how to write your conclusion. Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

FUTURE WORK

This section of the document points towards potential improvements, expansions, or new lines of inquiry that follow from your project. This section is crucial – not just for academic projects, but also for industry projects – as it acknowledges that research is an ongoing process and that one project can't cover every aspect of the topic.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

LITERATURE CITED

CRESCENZI, P., & KANN, V. (1997). Approximation on the web: A compendium of np optimization problems. In *Random'97* (p. 111-118).