**Scope Management Plan**

**XÂY DỰNG MÔ HÌNH PHÂN LOẠI VÀ TÓM TẮT VĂN BẢN VÀ HÌNH ẢNH**

# **OVERVIEW**

* 1. **Project background and description**

- Tên dự án: XÂY DỰNG MÔ HÌNH PHÂN LOẠI VÀ TÓM TẮT VĂN BẢN VÀ HÌNH ẢNH

- Trưởng dự án: Đoàn Văn Nguyên

- Chủ nhiệm dự án: Nguyễn Đức Anh

- Ngày bắt đầu dự án: 13 tháng 9 năm 2024

- Ngày kết thúc dự án: 27 tháng 12 năm 2024

- Kinh phí dự kiến: 10000$

- Phiên bản: 1

* + 1. **Statement**

Tóm tắt đa phương tiện đã thu hút nhiều sự chú ý do sự phát triển nhanh chóng của dữ liệu đa phương tiện. Đầu ra của các hệ thống tóm tắt đa phương tiện hiện tại thường được biểu diễn bằng văn bản. Tuy nhiên, việc chỉ dùng văn bản đang làm hạn chế thông tin thu thập được và khiến cho người dùng khó hình dung hơn. Nhận biết được hạn chế đó, trrong bài báo này, chúng tôi đề xuất một nhiệm vụ mới, tóm tắt đa phương tiện với đầu ra là cả hình ảnh và văn bản. Đầu vào gồm văn bản của 1 bài báo và 1 số các ảnh minh họa trong bài báo, đầu ra sẽ gồm : nội dung tóm tắt của bài báo đó và 1 hình ảnh có liên quan nhất tới nội dung tóm tắt đó.

* + 1. **Objective**
* đề xuất mô hình có kiến trúc mới so với các mô hình tóm tắt văn bản-hình ảnh hiện có
* Đào tạo và đánh giá các mô hình đã chọn để đảm bảo chúng đáp ứng các tiêu chuẩn về độ chính xác,
* mô hình cho kết quả vượt trội hơn các phương pháp tồn tại
* tóm tắt nội dung văn bản và chọn hình ảnh tương ứng, trên tập dữ liệu đầu vào lớn
* giải quyết được thách thức thách thức văn bản tóm tắt thiếu thông tin hình ảnh
* đạt hiệu xuất cao hơn trên các method hiện tại trên cả 2 bộ dữ liệu MNSO và CNN

**1.2 Project scope**

* **Scope includes:**
* độ dài file văn bản đầu vào không quá 20 trang
* định dạng file đầu vào: .txt với văn bản, .png với hình ảnh
* chỉ 1 hình ảnh phù hợp nhất với nội dung tóm tắt
* tạo bộ test cá nhân cho Tiếng Việt
* Phát triển các thuật toán để tóm tắt cả dữ liệu văn bản và hình ảnh.
* Tiến hành các thí nghiệm để tinh chỉnh các thông số mô hình và cải thiện hiệu suất
* **Scope does not includes:**
* văn bản định dạng pdf, epub,…, hình ảnh định dạng GIF, PSD,…
* tóm tắt cho đa hình ảnh
* bộ test cá nhân Tiếng Anh
* Phát triển các phương pháp cho các định dạng đa phương tiện khác (video, âm thanh)
* Xử lý dữ liệu thời gian thực cho các ứng dụng trực tiếp.,api,thực tế,…
* xây dựng giao diện

**1.3 Project Success Criteria**

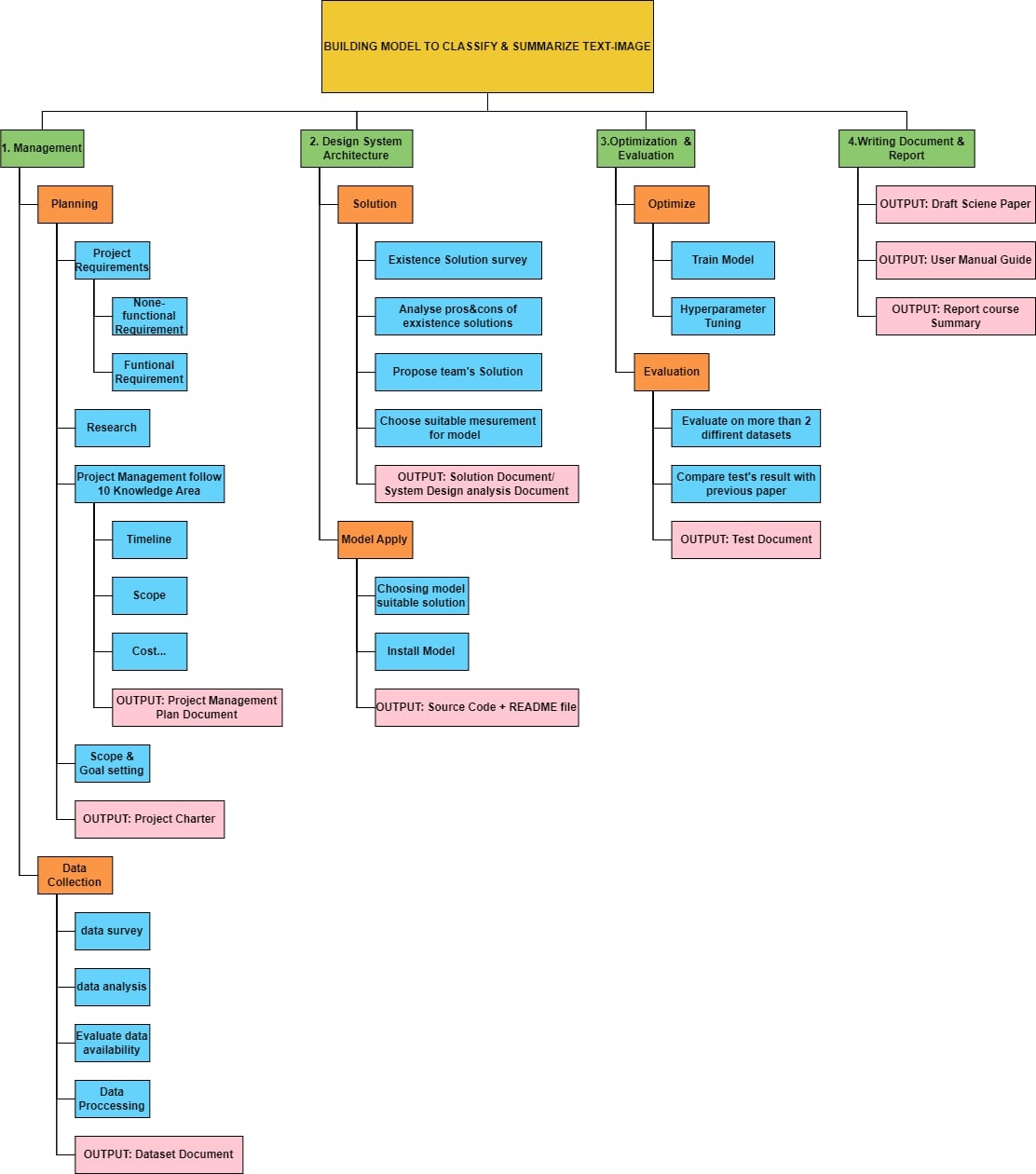
* **Quality:**
* .success criteria(Tiêu chí thành công) > 30%
* .ROUGE-1 F1 Score > 25%
* .đạt hiệu suất success criteria > 40% với tác vụ tóm tắt CHỈ văn bản trên bộ test cá nhân
* đạt hiệu suất success criteria > 35% với tác vụ tóm tắt CẢ văn bản và hình ảnh trên bộ test cá nhân
* **Cost:**
* Tổng chi phí của dự án không quá 10.000$
* **Time:**
* Dự án cần phải được hoàn thành và bàn giao đúng thời hạn như đã dự kiến ​​ban đầu
* **Scope:**
* Kết quả của dự án cần phải phù hợp với phạm vi ban đầu
* **Scalability:**
* Hệ thống phải có khả năng xử lý tập dữ liệu có dung lượng tối thiểu là 2GB bao gồm văn bản, chú thích và hình ảnh
* **Test Case:**
* Test trên ít nhất 2 bộ data lớn khác nhau

**1.4 Deliverables**

* **bài báo khoa học(draft science paper):** tổng hợp xu hướng, tập dữ liệu, mã dựa trên đó, giới thiệu + thử thách, kiến ​​trúc tổng quát, phương pháp, mô hình, kết quả, đánh giá
* **Báo cáo giải pháp(solution document ):** yêu cầu chức năng, yêu cầu phi chức năng, phần cứng, phần mềm của bên thứ ba, cấu hình và thiết kế
* **Bộ dữ liệu:** Miêu tả bộ dữ liệu
* **Mã nguồn & Hướng dẫn cài đặt(Source code+Installation document)**
* **Báo cáo kiểm thử(test document):** kế hoạch kiểm thử, chính sách kiểm thử, trường hợp kiểm thử, kịch bản kiểm thử, dữ liệu kiểm thử, báo cáo thực hiện kiểm thử, báo cáo lỗi
* **Tài Liệu Quản Lý Dự Án(Project Management Plan document):** tổng hợp theo 10 miền tri thức:(scope, time, cost,...)

# **WBS**

|  |
| --- |
| **Task Name** |
| **Build Model to Summarize text-image** |
| **1.Management** |
| **1.1 Planning** |
| 1.1.1 Research |
| 1.1.2 Project Requirements Document |
| 1.1.3 Project Management Plan Doc |
| 1.1.4 Project Charter |
| **1.2 Data Collection** |
| 1.2.1 Data Survey |
| 1.2.2 Data Analysis |
| **1.2.3 Data Processing** |
| 1.2.3.1 Data Normalization |
| 1.2.3.2 Data Cleaning |
| 1.2.3.3 Data Augmentation |
| 1.2.3.4 Data Analysis Document |
| **2. Design System Architecture** |
| **2.1 Choose Solution** |
| 2.1.1 Existence Solutions Survey |
| 2.1.2 Analyze Existence Solutions |
| 2.1.3 Propose Team's Solutions |
| 2.1.4 Choose Suitable Metrics |
| 2.1.5 Solutions/System Design Analysis Doc |
| **2.2 Model Apply** |
| 2.2.1 Choose Suitable Model |
| 2.2.2 Install Model |
| 2.2.3 Source Code + Readme File |
| **3. Optimization & Evaluation** |
| **3.1 Optimize** |
| 3.1.1 Train Model |
| 3.1.2 Hyperparameter Tuning |
| **3.2 Evaluation** |
| 3.2.1 Evaluate on 2 Datasets |
| 3.2.2 Compare Tests vs Other Papers |
| 3.2.3 Test Document |
| **4. Writing Document & Report** |
| 4.1 Draft Science Paper |
| 4.2 User Manual Guide |
| 4.3 Report Course Summary |



# **WBS Dictionary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1.Management** | Oversee the entire project, ensuring that each phase is planned, executed, and monitored effectively. | | | | |
| 1.1 Planning | Establish the roadmap and structure for the project. | |  |  |  |
| 1.1.1 Research | Conduct research to gather relevant information and insights. | | |  |  |
| 1.1.2 Project Requirements Document | Create a document outlining the project's objectives, needs, and constraints. | | | | |
| 1.1.3 Project Management Plan Doc | Draft a plan that details timelines, resources, and responsibilities for project execution. | | | | |
| 1.1.4 Project Charter | Define the scope, goals, and key stakeholders of the project. | | |  |  |
| **1.2 Data Collection** | Gather and process data to be used in the project. | |  |  |  |
| 1.2.1 Data Survey | Collect and review data from various sources. | |  |  |  |
| 1.2.2 Data Analysis | Examine the collected data to extract meaningful insights. | | |  |  |
| **1.2.3 Data Processing** | Prepare and clean the data for use in the project. | |  |  |  |
| 1.2.3.1 Data Normalization | Standardize the data to ensure consistency across the dataset. | | |  |  |
| 1.2.3.2 Data Cleaning | Remove or correct any errors or inconsistencies in the data. | | |  |  |
| 1.2.3.3 Data Augmentation | Enhance the data through various techniques to improve model performance. | | | | |
| 1.2.3.4 Data Analysis Document | Summarize the findings and insights from the data analysis. | | |  |  |
| **2. Design System Architecture** | Create the structure and design of the system to meet project goals. | | | |  |
| **2.1 Choose Solution** | Evaluate and select the best solution for the system's architecture. | | | |  |
| 2.1.1 Existence Solutions Survey | Review existing solutions to understand their effectiveness. | | |  |  |
| 2.1.2 Analyze Existence Solutions | Compare and assess the strengths and weaknesses of the current solutions. | | | |  |
| 2.1.3 Propose Team's Solutions | Present the solutions developed by the project team. | |  |  |  |
| 2.1.4 Choose Suitable Metrics | Select appropriate metrics to evaluate the system's performance. | | | |  |
| 2.1.5 Solutions/System Design Analysis Doc | Document the system's architecture and the rationale for design choices. | | | |  |
| **2.2 Model Apply** | Integrate the chosen model into the system. | |  |  |  |
| 2.2.1 Choose Suitable Model | Select the best model to apply to the project requirements. | | |  |  |
| 2.2.2 Install Model | Set up the chosen model in the system environment. | |  |  |  |
| 2.2.3 Source Code + Readme File | Provide the source code along with documentation to guide users. | | | |  |
| **3. Optimization& Evaluation** | Improve system performance and evaluate its success. | | |  |  |
| **3.1 Optimize** | Refine and enhance the model for better accuracy and efficiency. | | | |  |
| 3.1.1 Train Model | Train the model using the processed data. |  |  |  |  |
| 3.1.2 Hyperparameter Tuning | Adjust model parameters to optimize performance. | |  |  |  |
| **3.2 Evaluation** | Assess the system's performance and compare it with other benchmarks. | | | |  |
| 3.2.1 Evaluate on 2 Datasets | Test the model on two different datasets to measure its effectiveness. | | | |  |
| 3.2.2 Compare Tests vs Other Papers | Compare the test results with findings from similar studies. | | |  |  |
| 3.2.3 Test Document | Create a document detailing the testing procedures and outcomes. | | | |  |
| **4. Writing Document& Report** | Prepare final reports and documents for submission and use. | | |  |  |
| **4.1 Draft Science Paper** | Write a draft of the scientific paper summarizing the research. | | |  |  |
| **4.2 User Manual Guide** | Develop a guide to help users understand and operate the system. | | | |  |
| **4.3 Report Course Summary** | Write a comprehensive report summarizing the project's progress and outcomes. | | | | |

1. **SCHEDULE TIME-COST:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Task Name** | **Duration** | **Start** | **Finish** | **Cost1** | **Predecessors** |
| **Build Model to Summarize text-image** | **78 days?** | **Fri 9/13/24** | **Fri 12/27/24** | **$9,000.00** |  |
| **1.Management** | **25 days?** | **Fri 9/13/24** | **Tue 10/15/24** | **$2,200.00** |  |
| **1.1 Planning** | **12 days?** | **Fri 9/13/24** | **Thu 9/26/24** | **$500.00** |  |
| **1.1.1 Research** | **3 days?** | **Fri 9/13/24** | **Sun 9/15/24** | **$100.00** |  |
| **1.1.2 Project Requirements Document** | **3 days?** | **Mon 9/16/24** | **Wed 9/18/24** | **$100.00** | **4** |
| **1.1.3 Project Management Plan Doc** | **3 days?** | **Thu 9/19/24** | **Mon 9/23/24** | **$150.00** | **5** |
| **1.1.4 Project Charter** | **3 days?** | **Tue 9/24/24** | **Thu 9/26/24** | **$150.00** | **6** |
| **1.2 Data Collection** | **15 days?** | **Wed 9/25/24** | **Tue 10/15/24** | **$1,700.00** |  |
| **1.2.1 Data Survey** | **3 days?** | **Wed 9/25/24** | **Fri 9/27/24** | **$300.00** |  |
| **1.2.2 Data Analysis** | **2 days?** | **Mon 9/30/24** | **Tue 10/1/24** | **$300.00** | **9** |
| **1.2.3 Data Processing** | **11 days?** | **Tue 10/1/24** | **Tue 10/15/24** | **$1,100.00** |  |
| **1.2.3.1 Data Normalization** | **2 days** | **Fri 10/4/24** | **Sun 10/6/24** | **$300.00** |  |
| **1.2.3.2 Data Cleaning** | **3 days** | **Mon 10/7/24** | **Wed 10/9/24** | **$300.00** | **12** |
| **1.2.3.3 Data Augmentation** | **3 days** | **Thu 10/10/24** | **Sat 10/12/24** | **$300.00** | **13** |
| **1.2.3.4 Data Analysis Document** | **3 days?** | **Wed 10/2/24** | **Fri 10/4/24** | **$200.00** | **10** |
| **2. Design System Architecture** | **33 days?** | **Wed 10/16/24** | **Fri 11/29/24** | **$2,100.00** |  |
| **2.1 Choose Solution** | **19 days?** | **Wed 10/16/24** | **Mon 11/11/24** | **$1,050.00** |  |
| **2.1.1 Existence Solutions Survey** | **3 days?** | **Mon 10/21/24** | **Wed 10/23/24** | **$250.00** |  |
| **2.1.2 Analyze Existence Solutions** | **3 days?** | **Thu 10/24/24** | **Sat 10/26/24** | **$250.00** | **18** |
| **2.1.3 Propose Team's Solutions** | **3 days?** | **Mon 10/28/24** | **Wed 10/30/24** | **$250.00** | **19** |
| **2.1.4 Choose Suitable Metrics** | **3 days** | **Thu 10/31/24** | **Mon 11/4/24** | **$250.00** | **20** |
| **2.1.5 Solutions/System Design Analysis Doc** | **5 days** | **Tue 11/5/24** | **Mon 11/11/24** | **$50.00** | **21** |
| **2.2 Model Apply** | **17 days** | **Thu 11/7/24** | **Fri 11/29/24** | **$1,050.00** |  |
| **2.2.1 Choose Suitable Model** | **3 days** | **Tue 11/12/24** | **Thu 11/14/24** | **$500.00** | **22** |
| **2.2.2 Install Model** | **9 days** | **Fri 11/15/24** | **Wed 11/27/24** | **$500.00** | **24** |
| **2.2.3 Source Code + Readme File** | **2 days** | **Thu 11/28/24** | **Fri 11/29/24** | **$50.00** | **25** |
| **3. Optimization& Evaluation** | **18 days?** | **Sat 11/23/24** | **Tue 12/17/24** | **$3,800.00** |  |
| **3.1 Optimize** | **13 days** | **Sat 11/23/24** | **Tue 12/10/24** | **$2,500.00** |  |
| **3.1.1 Train Model** | **2 days** | **Thu 11/28/24** | **Fri 11/29/24** | **$500.00** | **25** |
| **3.1.2 Hyperparameter Tuning** | **7 days** | **Mon 12/2/24** | **Tue 12/10/24** | **$2,000.00** | **29** |
| **3.2 Evaluation** | **6 days** | **Tue 12/10/24** | **Tue 12/17/24** | **$1,300.00** |  |
| **3.2.1 Evaluate on 2 Datasets** | **4 days** | **Mon 12/2/24** | **Thu 12/5/24** | **$750.00** | **29** |
| **3.2.2 Compare Tests vs Other Papers** | **2 days** | **Wed 12/11/24** | **Thu 12/12/24** | **$50.00** | **29,30** |
| **3.2.3 Test Document** | **3 days** | **Sun 12/15/24** | **Tue 12/17/24** | **$500.00** |  |
| **4. Writing Document& Report** | **16 days** | **Sun 12/1/24** | **Fri 12/20/24** | **$900.00** |  |
| **4.1 Draft Science Paper** | **17 days** | **Thu 11/28/24** | **Fri 12/20/24** | **$300.00** | **25** |
| **4.2 User Manual Guide** | **5 days** | **Thu 11/28/24** | **Wed 12/4/24** | **$300.00** | **25** |
| **4.3 Report Course Summary** | **9 days** | **Thu 11/28/24** | **Tue 12/10/24** | **$300.00** | **25** |