Table

Description automatically generated with medium confidence

****

**TABLE OF CONTENTS**

[**1.** **EXECUTIVE SUMMARY** 5](#_Toc105668736)

[1.1 FINANCIAL PROPOSAL 6](#_Toc105668737)

[1.2 BROAD-BASED BLACK ECONOMIC EMPOWERMENT 7](#_Toc105668738)

[**2.** **BIDDER’S RESPONSE** 8](#_Toc105668739)

[**3.** **ADVANTAGES OF USING NIGHTCORE MECH** 11](#_Toc105668740)

[**4.** **BACKGROUND** 12](#_Toc105668741)

[**5.** **PROJECT TEAM** 13](#_Toc105668742)

[**6.** **UNDERSTANDING OF PROJECT BRIEF** 15](#_Toc105668743)

[**7.** **AIMS AND OBJECTIVES** 17](#_Toc105668744)

[**8.** **DELIVERABLES** 18](#_Toc105668745)

[**9.** **METHODOLOGIES** 19](#_Toc105668746)

[9.1 Design Methodology to be utilised 19](#_Toc105668747)

[9.2 A few reasons to use the Agile-Waterfall Hybrid model: 19](#_Toc105668748)

[9.3 Pros of a hybrid methodology model: 19](#_Toc105668749)

[9.4 Project Management Methodology to be utilised 19](#_Toc105668750)

[**10.** **PROJECT PLAN** 20](#_Toc105668751)

[**11.** **BUDGET** 20](#_Toc105668752)

[**12.** **INVOICING SCHEDULE** 23](#_Toc105668753)

[11.1 12.1. Terms and Conditions 24](#_Toc105668754)

[11.2 12.2. Purpose and Context 24](#_Toc105668755)

[11.3 12.3. Validity of Submission 24](#_Toc105668756)

[11.4 12.4. Contractual Relationship 24](#_Toc105668757)

[11.5 12.5. Duration 24](#_Toc105668758)

[11.6 12.6. Client’s Responsibility 25](#_Toc105668759)

[11.7 12.7. Invoicing and payment 25](#_Toc105668760)

[11.8 12.8. Risk 25](#_Toc105668761)

[11.9 12.9. Confidentiality and publication 25](#_Toc105668762)

[11.10 12.10. Liability 26](#_Toc105668763)

[11.11 12.11. Intellectual property 26](#_Toc105668764)

[11.12 12.12. Breach 26](#_Toc105668765)

[11.13 12.13. Termination 26](#_Toc105668766)

[11.14 12.13. Waiver 26](#_Toc105668767)

[11.15 12.14. Total Agreement and Amendments 27](#_Toc105668768)

[11.16 12.15. Governing Law 27](#_Toc105668769)

[11.17 12.16. Domicilia Citande et Exencutandi and Notice 27](#_Toc105668770)

[11.18 12.17. Precedence 27](#_Toc105668771)

[**13.** **RELEVANT EXPERIENCE** 27](#_Toc105668772)

[**14.** **CONTACT DETAILS OF BIDDER** 28](#_Toc105668773)

[**15** **ANNEXURE A - CURRICULUM VITAE OF PROJECT MANAGEMENT TEAM** 29](#_Toc105668774)

[**16** **ANNEXURE B – PROJECT MANAGEMENT** 30](#_Toc105668775)

[**1.** **Work Breakdown Structure** 30](#_Toc105668776)

[1.1. Phases 30](#_Toc105668777)

[1.2. Activities 30](#_Toc105668778)

[**1.2.1** **Initiation Phase** 30](#_Toc105668779)

[**1.2.2** **Planning Phase** 30](#_Toc105668780)

[**1.2.3** **Execution Phase** 30](#_Toc105668781)

[**1.2.4** **Project Review/Project Closure Phase** 30](#_Toc105668782)

[**1.2.5** **Maintenance Phase** 30](#_Toc105668783)

[1.3 Tasks 30](#_Toc105668784)

[**1.3.1** **Initiation Phase** 30](#_Toc105668785)

[**1.3.2** **Planning Phase** 30](#_Toc105668786)

[**1.3.3** **Execution Phase** 30](#_Toc105668787)

[**1.3.4** **Project Review/Project Closure Phase** 30](#_Toc105668788)

[**1.3.5** **Maintenance Phase** 30](#_Toc105668789)

[1.4 Milestones 30](#_Toc105668790)

[1.5 Effort 30](#_Toc105668791)

[**2.** **Project Plan** 30](#_Toc105668792)

[2.1 Schedule 30](#_Toc105668793)

[2.2 Dependencies 30](#_Toc105668794)

[2.3 Assumptions 30](#_Toc105668795)

[2.4 Constraints 31](#_Toc105668796)

[**3.** **Project Scope and Goals** 31](#_Toc105668797)

[**4.** **Deliverables** 32](#_Toc105668798)

[4.1 Documentation Oriented Deliverables 32](#_Toc105668799)

[4.2 Secondary Documentation Deliverables 32](#_Toc105668800)

[4.3 Primary Deliverable 32](#_Toc105668801)

[4.4 Secondary Deliverables 32](#_Toc105668802)

[**4.4.1** **User Manuals** 32](#_Toc105668803)

[**4.4.2** **User Training** 32](#_Toc105668804)

[4.5 Stakeholders 32](#_Toc105668805)

[**5.** **Critical Success Factors** 32](#_Toc105668806)

[**6.** **Risk Plan** 33](#_Toc105668807)

[**7.** **Project Plan Gantt Chart** 35](#_Toc105668808)

[**8.** **Quality Plan** 35](#_Toc105668809)

[**9.** **Financial Plan** 35](#_Toc105668810)

[9.1 Labour 35](#_Toc105668811)

[9.2 Equipment 35](#_Toc105668812)

[9.3 Schedule 35](#_Toc105668813)

[9.4 Assumptions 35](#_Toc105668814)

[9.5 Financial Processes 35](#_Toc105668815)

[9.6 Activities 35](#_Toc105668816)

[9.7 Roles 35](#_Toc105668817)

[9.8 Documents 35](#_Toc105668818)

[**10.** **Procurement Plan** 35](#_Toc105668819)

[10.1 Procurement Requirements 35](#_Toc105668820)

[10.2 Requirements 35](#_Toc105668821)

[10.3 Market Research 35](#_Toc105668822)

[10.4 Schedule 35](#_Toc105668823)

[10.5 Assumptions 35](#_Toc105668824)

[10.6 Constraints 35](#_Toc105668825)

[10.7 Activities 35](#_Toc105668826)

[10.8 Roles 35](#_Toc105668827)

[**11.** **Resource Plan** 36](#_Toc105668828)

[11.1 Labour 36](#_Toc105668829)

[11.2 Equipment 36](#_Toc105668830)

[11.3 Assumptions 36](#_Toc105668831)

[11.4 Constraints 36](#_Toc105668832)

[**12.** **Communication Plan** 36](#_Toc105668833)

[12.1 Stakeholder List 36](#_Toc105668834)

[12.2 Stakeholder Requirements 36](#_Toc105668835)

[12.3 Assumptions 36](#_Toc105668836)

[12.4 Constraints 36](#_Toc105668837)

[12.5 Activities 36](#_Toc105668838)

[12.6 Roles 36](#_Toc105668839)

[12.7 Methods, Documents and Technology 36](#_Toc105668840)

[**13.** **Appendix C: Nightcore Mech Meeting Minutes** 36](#_Toc105668841)

# **EXECUTIVE SUMMARY**

NightCore Mech is a company that provides software solutions to enterprises and build systems that provide value to business processes and analytics. The company has been around since 2019 and is located in Centurion East, Gauteng. We provide our clients with efficient solutions that aid in the growth and profitability of their enterprises. NightCore Mech consists of project managers, frontend and backend developers, SQL developers and database administrators, as well as prototyping developers and unit testers. Each member of the company has a variety of skills that contributes to knowing how to implement and design systems to address business requirements. Our project team is experienced and confident in our capabilities to perform a feasibility study on the design and implementation of software to identify academic misconduct committed by students at the NWU. The current requirements from the client (NWU Registrar), with Mr Zander Janse van Rensburg as the project overseeing manager, requires our company to design and build a modular workflow system that would assist lecturers in academics to identify and report academic misconduct cases according to standing NWU SOPS. The NWU Registrar must address plagiarism by evaluating each case individually and appointing experts to prepare technical reports. External subject matter experts (SMEs) are requested to examine the technical reports with an additional report that provides a deeper insight on the alleged plagiarism if the technical reports do not self-evidently emphasize the severity of the plagiarism. Manually comparing the allegedly plagiarized text in issue with the original text as evidence text is a requirement for the technical report, which can become difficult and lead to certain similarities being overlooked. The developers of NightCore Mech must adhere to standard project management methodologies and frameworks/bodies to plan and implement the lifecycle of this project, with a hybrid methodology between the Waterfall and Agile models or the suggested MPMM / Method 123 that will be implemented. The project team conducted some recent research on project management methodologies/frameworks and concluded that about 70% of projects fail due to time and budget constraints, as well as scope requirements and changes.

According to the requests of the bidder, the academic misconduct system should adhere to the following:

* The automation of comparisons of question text with evidence text and it must produce a similarity measure for the two texts (2019-2021). Through the use of a similarity metric, the software should reduce the time spent manually comparing two texts and generalizing the assessment of how severe the conjectured copying is. The software must combine text-matching skills with stylometric analytics to provide more accurate reports, better explain academic misbehaviour, and enable improved decision-making.
* The current software offers capabilities like a crude text comparison tool and a stylometric tool that does not perform optimally. The rudimentary text comparison tool is used to point out similarities between texts in technical reports, whereas the stylometric tool is underutilized due to its extensive possibilities in assisting the investigator in authorship recognition.
* The application should be standalone and terrestrial-based, being able to operate on multiple systems at once.
* Daily data updates should occur.
* The system should be able to maintain database transactions over time.
* The development method/methodology must adhere to a proper project management methodology. The MPMM / Method 123 or a hybrid Waterfall/Agile method is preferred.
* The project should conform to ISO 21500 standards firstly, secondly to PMBoK as well as the PRINCE2 Method and the HERMES method.
* The application should be able to detect text similarities between two sources as a minimum requirement.
* The application should have stylometric functionalities, the application should be able to detect original authorship of a document.
* Application should be memory-efficient and resource overhead should be avoided.
* Feedback reports should be based on stylometric performance metrics used to identify authorship and contract-cheating.
* Statistical analysis and knowledge should be crucial during the development phase.
* Instead of creating a database to persist data, the development team can create a local corpus of documents to compare and identify contract-cheating.

Based on the abovementioned requirements, the development team of NightCore Mech decided to develop a web-based application that facilitates ease of access and efficient integrations should be developed. The web application can be hosted on efficient cloud platforms such as Microsoft Azure or AWS. If it happens that hosting with the aforementioned platforms are not a feasible option, the development team will make the necessary arrangements with the NWU Registrar to host the application.

## FINANCIAL PROPOSAL

The NightCore Mech contractors will deliver all the requirements as requested by the client within 8 months at a total cost of R852 997.37 (VAT and disbursements included).

## BROAD-BASED BLACK ECONOMIC EMPOWERMENT

NightCore Mech is considered a Level 2 contributor with a 0% B-BBEE procurement level. Because the company earned less than R10 000 000 in the last financial year, the company qualifies for the B-BBEE Exempted Micro Enterprise BEE exemption.

# **BIDDER’S RESPONSE**

|  |  |
| --- | --- |
| **Criteria for Functional Evaluation** | **Research Solutions’ Response** |
| Details about the company's software development projects that the team has completed in the last 12 months. These initiatives will be reviewed, with information about the project sponsors provided in the section below. | **Entire team**  Completed a Vitality Web App that manages Active Rewards.  Completed a Social Media Web application in which users could share photos.  Completed a database system for a fire damage assessment system. |
| Detailed CVs of team members who will be directly involved in delivering the required services. The individuals must have prior experience in software development. | Every member’s full CV can be found in the addendums section. The following is a summary of their experience:  **Hanno Visagie**  3 years of coding experience in C#, C++, Java, CSS, HTML, SQL and ASP.NET. 3 years of experience in project management along with experience in systems analysis and design. 3 years of experience in decision support systems and business intelligence along with data analysis.  **Hano Strydom**  1 Years’ experience using ASP.Net, Maven, Spring  2 Years’ experience using Oracle & SQL  Worked on numerous projects  3 Years’ experience in programming (Python, Java, C#)  **Michael Rosin**  3 years of web design (back-end) using ASP.NET, C#, .NET and database management using Oracle, SQL Server, MS Access, MS SQL and implementing BI/ETL solution. 3 years of programming experience such as C#, C, C++, Python, Java, VB.NET and PHP.  **Annika du Toit**  3 - years information technology experience which includes coding in several languages. (Java, Python, C\C++, C#, ASP.NET).  2-year experience in database management (ORACLE).  **Llewelyn Anthony**  3 years of experience developing .NET desktop applications and ASP.NET MVC applications.  2 years of experience working with HTML5/CSS3/JavaScript and React Single-Page applications while using Git as a source-control tool.  5 years of experience as a SQL developer and a year of experience with Postgres.  A year’s experience worth of Python and NodeJS scripting.  Recent training on setting up Docker containers and execs.  Spent 2 years as a System analysis and software lifecycle mentor.  **Shené Boshoff**  3 years of programming and software development experience.  2 years of database management experience.  1 year of systems analysis experience. |
| A minimum of three (3) contactable references where the bidder has completed the software project. | Janet Liebenberg  Senior Lecturer  North-West University  Email: Janet.Liebenberg@nwu.ac.za  Tel No: +27 18 299 2536  Linda Redelinghuys  Lecturer  North-West University  Email: 10143882@nwu.ac.za  Tel No: +27 18 299 2531  Zander Boonzaaier  Lecturer  North-West University  Email: 28749995@student.g.nwu.ac.za |
| Based on the deliverables referred to in Section 5 above, a detailed project plan with timelines on how the assignment will be carried out including the risks that might be involved (including risk mitigation factors). | A full project plan and budget can be found in sections 10 and 11. The project will take place over 9 months with 8 phases. The phases are the following:   1. Scope definition and project planning 2. Problem analysis 3. Requirement analysis 4. Logical design 5. Decision analysis 6. Physical design and implementation 7. Testing 8. Installing and deployment |

# **ADVANTAGES OF USING NIGHTCORE MECH**

NightCore Mech is a respected private company that has learned and grown exponentially in the past three years. We happen to be a dream team of marketers, project managers, sales leaders, web designers and end-to-end eCommerce consultants who are experts in our industries. We specialize in converting Enterprise eCommerce business concepts into scalable, high-performing storefronts with robust, functional, and user-friendly solutions. We provide unique software solutions to organizations and individuals. Based on our own experience, we provide expert opinions, analysis, and suggestions to organizations or individuals. We are expert problem-solvers who offer solutions to overcome enterprise obstacles and improve performance.

The key advantages of hiring us as your e-commerce software consultant are as follows:

1. Perspective:

Simply receiving an outsider’s perspective on your company can aid in its growth. Our e-commerce software experts work with a variety of businesses, observing which ideas and approaches are feasible. Any company can become trapped in an echo chamber, which stifles innovation. We can assist you in re-imagining your business operations and marketing tactics, in addition to making software recommendations and managing e-commerce optimization on your site.

1. Improved customer experience:

We notice difficulties and propose solutions.

1. Streamlined workflow:

We can assess your company procedures and software, drawing on years of experience assisting other e-commerce businesses, to offer relevant solutions. Identifying areas of your company that could benefit from process automation can help you boost your profile.

1. A competitive edge:

We use an omnichannel strategy to e-commerce to help your company obtain or keep a competitive advantage. However, if your software isn’t integrated across channels, you risk missing out on sales or keeping clients waiting too long for a solution to arrive.

1. Branch out into new sales channels:

To generate visitors to your website, e-commerce success necessitates more than just SEO or paid search. You can utilize omnichannel marketing to reach out to customers in a variety of ways.

1. Expert evaluation:

Without re-platforming, an e-commerce software expert investigates numerous options for improving your site’s performance and functionality. Our professional e-commerce consultants concentrate on iterative, platform-agnostic enhancements to your applications. We won’t offer you a solution redesign or features you don’t need unless they provide value to your company. However, if your e-commerce software is at the end of its life cycle, an e-commerce expert can assist you with the re-platforming process.

# **BACKGROUND**

The Registrar of the North-West University is responsible for various administration tasks within the university. These tasks include record keeping of university student marks and a wide range of other supporting records and documents. The university’s view on information governance, in conjunction with Gartner’s definition, states that it views information governance as an overarching framework that provides an oversight of information and the processes by which it is generated, processed, and curated at the university.

The NWU Registrar is responsible for investigating various cases of alleged plagiarism committed by students, by inspecting cases individually and appointing experts to develop technical reports. The technical report currently involves manually comparing the alleged plagiarised text in question with the original text as evidence text, which can be tedious, and sometimes leads to similarities being overlooked.

If the technical reports fail to emphasise the severity of the alleged plagiarism, external subject matter experts may be requested to adjudicate the technical report with an additional report which further elaborates on the alleged plagiarism.

This tender, therefore calls for the design and development of a workflow system that will be able to facilitate academics in reporting academic misconduct cases according to standing NWU SOPs.

# **PROJECT TEAM**

|  |  |
| --- | --- |
|  | **Project Leader**  Hanno Visagie  Qualifications:   * BSc Information Technology (North-West University) * Current BSc (Hons) Computer Science and Information Systems student at North-West University |
| A person sitting on a boat  Description automatically generated with medium confidence | **Prototype Lead**  Hano Strydom  Qualifications:   * BSc Information Technology (North-West University) * Current BSc (Hons) Computer Science and Information Systems student at North-West University |
| A person standing in front of a body of water  Description automatically generated with medium confidence | **Back-end Lead**  Michael Rosin  Qualifications:   * BSc Information Technology (North-West University) * Current BSc (Hons) Computer Science and Information Systems student at North-West University |
|  | **Database Administrator**  Annika du Toit  Qualifications:   * BSc Information Technology (North-West University) * Current BSc (Hons) Computer Science and Information Systems student at North-West University |
| A person smiling for the camera  Description automatically generated with medium confidence | **Development Lead**  Llewellyn Anthony  Qualifications:   * BSc Information Technology (North-West University) * Current BSc (Hons) Computer Science and Information Systems student at North-West University * 2017 – 2018 QCTO National Certification Information Technology: Systems Development (SAQA ID 48872, NQF 5, 131 Credits). MCSD and BI. |
| A person smiling for the camera  Description automatically generated with medium confidence | **Hosting Platform Manager**  Shené Boshoff  Qualifications:   * BSc Information Technology (North-West University) * Current BSc (Hons) Computer Science and Information Systems student at North-West University |

# **UNDERSTANDING OF PROJECT BRIEF**

NightCore Mech will aim to integrate a software-based solution that will address the following business requirements:

* Design and develop a standalone terrestrial-based system that can be used by multiple systems at a time.
* The integration and use of project management method123 for designing and developing the project, or an alternative hybrid Waterfall/Agile method.
* Promote the significance of a simple and user-friendly interface.
* Eliminate unnecessary time wasted while using the software
* Increase the efficiency of staff members by delivering software that works.
* Compile documentation/user manuals for employees, in a category-based framework, which will act as a guide to avoid confusion.
* Work closely with stakeholders to ensure that all the requirements are met.
* Ensure concurrent use of the software.
* At the end of the project lifecycle, the developers at Nightcore Mech should be able to deliver a system that enables lecturers at the NWU to identify academic misconducts/contract-cheating.
* Feedback Reports should be provided by the system indicating text similarities between various sources to identify plagiarism on different levels conducted by a student.
* Application should be web-based
* Automation of feedback reports, with stylometric performance metrics visualised in order to detect contract-cheating and original authorship
* Side-by-side comparison of source document and evidence texts, without the need to manually identify text similarities
* Machine learning algorithms and stylometric analysis should be implemented to detect authorship and text similarities
* Text similarities is a minimum requirement of the application
* Use document metadata to support stylometric analysis
* A local corpus should be stored on a device to compare documents and identify contract-cheating and authorship
* Computing resources should be managed efficiently, and memory overhead should be avoided
* Daily data updates should occur.
* The system should be able to maintain database transactions over time.

The NightCore Mech team will revise the previous project to obtain some valuable insight into understanding what the exact requirements of the stakeholders are and what fell short of their expectations. Software solutions often fall short of the stakeholder expectations due to not satisfying the business requirements as a result of poor communication or failure to adhere to system design methodologies or best practices. To overcome these challenges, NightCore Mech utilises a holistic approach and strictly follows various methodologies that facilitate communication and collaboration throughout the various phases and environments of the project. Please refer to Figure 1 for more information



**Figure 1: Environments in the holistic approach to information and communications technology management**

Subsequently, NightCore Mech will follow this holistic approach to ensure that the best decisions regarding the development of software are chosen that will satisfy the requirements provided by stakeholders. To achieve this goal, NightCore Mech must identify which aspects of the software solution are of high importance by conducting an all-inclusive feasibility study that will identify crucial aspects and requirements in conjunction with the objectives established by stakeholders.

# **AIMS AND OBJECTIVES**

Nightcore mech facilitate academics in REPORTING ACADEMIC MISCONDUCT CASES ACCORDING TO STANDING NWU SOPS.

A picture containing text, ax, tool, gear

Description automatically generated

**Objective 1: Comply to all relevant legislations and regulations indicated. This include but are not limited to the Legal Ramifications of Plagiarism, The Copyright Act, The NWU policy on plagiarism and other forms of academic misconduct.**

A picture containing text, ax, tool, gear

Description automatically generated

**Objective 2: Improve the facilitation of academics in reporting academic misconduct cases within the NWU by monitoring the level of plagiarism committed by students.**

A picture containing text, ax, tool, gear

Description automatically generated

**Objective 3: Upgrade current technological application to enhance the ability of the NWU to provide more accurate plagiarism reports.**

A picture containing text, ax, tool, gear

Description automatically generated

**Objective 4: Introduce a technological solution that includes:**

* **Software that detects numerous ways of paraphrasing and automatically compare two similar texts.**
* **Understand academic misconduct better and provide an accurate report for each student to determine those suspected from cheating.**
* **Executable standalone terrestrial-based system working on multiple platforms.**
* **Have the capacity to update data on a daily basis.**
* **Database maintenance of all transactions over time.**
* **Adhere to a proper PM methodology.**

# **DELIVERABLES**

Nightcore mech facilitate academics in REPORTING ACADEMIC MISCONDUCT CASES ACCORDING TO STANDING NWU SOPS.

A picture containing text, ax, tool, gear

Description automatically generated

**Deliverable 1: Include project purpose and scope, develop project plan, specify qualifications of team, determine roles and responsibilities, risk report, progress reports.**

A picture containing text, ax, tool, gear

Description automatically generated

**Deliverable 2: Budget Analysis of proposed project solutions and requirements.**

A picture containing text, ax, tool, gear

Description automatically generated

**Deliverable 3: Implement the proposed solution and report performance outcomes of technological solutions**

A picture containing ax, tool

Description automatically generated

**Deliverable 4: A document that lists all the technical requirements – this include but are not limited to:**

* **Software architecture design document**
* **User requirement document**
* **User manual and user training for technological solution**
* **Ensure that the customer owns all source codes, database scripts, content, and associated documentation.**

A picture containing ax, tool

Description automatically generated

**Deliverable 5: Management and the Board will be provided with a project completion report.**

# **METHODOLOGIES**

## Design Methodology to be utilised

NightCore Mech utilises a combination of Agile practices and the Waterfall software development methodology to effectively and efficiently develop software that satisfies all the client’s needs. The combined use of Agile and Waterfall will shorten the delivery time and provide the company with the ability to gather feedback during the early stages of development to ensure that the requirements are met successfully.

## A few reasons to use the Agile-Waterfall Hybrid model:

* Creates an environment for enhanced collaboration
* Decreases the time required for design, analysis and planning
* Clearly defines various project frames such as budget and time delivery
* Ensures software follows software standards

## Pros of a hybrid methodology model:

* Deliverables are agreed on by the client and developers at an early stage in the development lifecycle
* Easily track project progression as the entire scope of work is known in advance
* Encourages the project to be divided into smaller delivery segments, referred to as sprints, to allow easier development of a complex system and to provide an overview of progression.

The successful use and implementation of a hybrid design methodology will ensure a smooth design and development during the different stages of the development lifecycle.

## Project Management Methodology to be utilised

NightCore Mech utilises the project management methodology known as MPMM, which stands for Method123 Project Management Methodology. The use of this methodology is to ensure that best practise and standards are met in the workplace.

PMBOK, or Project Management Body of Knowledge, is a compilation of standard terminology and guidelines for project management that is utilised along with MPMM to further support the development lifecycle. MPMM and PBMOK integrate seamlessly as both methodologies attempt to utilise best practices.

PMBOK consists of five different process groups namely: Initiating, planning, execution, monitoring, controlling and closing. These process groups are integrated with each of the different elements of MPMM to boost the effectiveness of both methodologies.

Along with these two methodologies, NightCore Mech makes use of the ISO standard 21500:2012, which guides project management. ISO 21500:2012 provides high-level descriptions of concepts and processes which forms part of good practices in project management.

# **PROJECT PLAN**

The link below contains the project plan which has been done in excel.

<https://mega.nz/file/h1NwzDCT#tz5isgOkc2Xc3WYHvj1JegXXILFUO7RuaodZeoyGs_w>

# **BUDGET**

\*Any documents/page cost covered in miscellaneous

\*Presentations are virtually presented, held on Zoom.

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Work Component** | | **Amount** |
| **1** | **Phase 1: Scope definition and Project planning** | |  |
| 1.1 | Project kick-off meeting | | R0.00 |
| 1.2 | Approach and methodology discussions | | R0.00 |
| 1.3 | Stakeholder meetings and member roles | | R0.00 |
| 1.4 | Project proposal and official scope | | R0.00 |
| 1.5 | Project Schedule | | R2 775.60 |
| 1.6 | Presentation of phase 1 | | R564.20 |
|  | **TOTAL** | | **R3339.80** |
| **2** | **Phase 2: Problem Analysis** | |  |
| 2.1 | Problem and goal documentation | | R0.00 |
| 2.2 | Reference Framework | | R0.00 |
| 2.3 | SWOT analysis | | R0.00 |
| 2.4 | Technical requirements | | R0.00 |
| 2.5 | API research | | R0.00 |
| 2.6 | Presentation of phase 2 | | R564.20 |
|  | **TOTAL** | | **R564.20** |
| **3** | **Phase 3: Requirements Analysis** | |  |
| 3.1 | Documenting and analysing requirements | | R0.00 |
| 3.2 | Functional and non-functional requirements | | R0.00 |
| 3.3 | Interview with stakeholder (Travel Costs) | | R4 800.00 |
| 3.4 | Feasibility study | | R140 000 |
| 3.5 | Presentation of phase 3 | | R564.20 |
|  | **TOTAL** | | **R145 364.20** |
| **4** | **Phase 4: Logical Design** | |  |
| 4.1 | Design conceptual model | | R0.00 |
| 4.2 | Map conceptual model to logical model | | R0.00 |
| 4.5 | Prototyping of model | | R15 046 |
| 4.4 | Validate logical model | | R0.00 |
| 4.5 | Presentation of prototype | | R4 800.00 |
| 4.6 | Presentation of phase 4 | | R564.20 |
|  | **TOTAL** | | **R20 410.20** |
| **5** | **Phase 5: Decision Analysis** | |  |
| 5.1 | Documentation of objectives and evaluation criteria | | R0.00 |
| 5.2 | Alternative methodologies and design research | | R0.00 |
| 5.3 | Risk assessment | | R0.00 |
| 5.4 | Evaluation and presentation of risk assessment | | R0.00 |
| 5.5 | Presentation of phase 5 | | R564.20 |
|  | **TOTAL** | | **R564.20** |
| **6** | **Phase 6: Physical design and implementation** | |  |
| 6.1 | Visual Studio Community | | R0.00 |
| 6.2 | SQL Server 2019 and SSMS 18 | | R0.00 |
| 6.3 | API ([plagiarismcheck.org](https://plagiarismcheck.org/)) | | R1800.00 |
| 6.4 | Hosting platform (Microsoft Azure / AWS) | | R84 000.00 |
| 6.6 | Network infrastructure (Stable Internet connection) | | R0.00 |
| 6.7 | Database design and implementation | | R0.00 |
| 6.8 | Application design (GUI and Back-end) | | R0.00 |
| 6.9 | Application Prototyping | | R0.00 |
| 6.10 | Demonstration of prototype | | R2 400.00 |
| 6.11 | Presentation of Phase 6 | | R564.20 |
|  | **TOTAL** | | **R88 764.20** |
|  | **Phase 7: Testing** | |  |
| 7.1 | Unit testing | | R0.00 |
| 7.2 | Building and testing of databases | | R0.00 |
| 7.2 | Presentation of Phase 7 | | R564.20 |
|  | **TOTAL** | | **R564.20** |
| **8** | **Phase 8: Installation and Deployment** | |  |
| 8.1 | Hardware:   * Computer   + i3 12100 GT 1030 Workstation   + ASUS PRIME B660M-K MB   + i3 12100 4x Cores CPU   + KLEVV CRAS XR 16GB 3600MhZ   + 512GB NVME SSD   + GeForce GT 1030 2GB GDDR5   + Standard Black case   + Standard 400W PSU   + 7.1 CH HD Sound * Monitor * Peripherals   + Keyboard   + Mouse   + Webcam   + Mic * UPS for workstation and Router | | R18 998  R5 998.00  R1 500.00  R4 998.00 |
| 8.2 | Final Application and implementation | | R0.00 |
| 8.3 | User Manual | | R0.00 |
| 8.4 | Application Maintenance | | R59 285.88 |
| 8.5 | Presentation of Phase 8 | | R564.20 |
|  | **TOTAL** | | **R91 344.10** |
|  | **SUB-TOTAL** | |  |
| **Hours** | **Contractor wage / tariff** | **R/Hour** | **Total** |
|  | Ricus Warmenhoven  - Project Manager | R1600.00 | R80 000.00 |
|  | Hanno Visagie - Chief Technical Officer  - Project Leader | R1666.00  R300.00 | R58 520.00  R13 500.00 |
|  | Hano Strydom - Front-end  - Prototype Lead | R1000.00  R600 | R30 000.00  R12 000.00 |
|  | Llewellyn Anthony  - Lead Developer  - Unit Testing | R924.00  R782.00 | R36 960.00  R15 640 |
|  | Michael Rosin - Back-end Lead  - Software Analyst | R1800.00  R440.00 | R72 000.00  R15 400.00 |
|  | Annika du Toit - SQL Developer  - Database Admin | R800.00  R600.00 | R24 000.00  R24 000.00 |
|  | Shené Boshoff - Back-end Developer  - Hosting Platform manager | R1400.00  R960.00 | R49 000.00  R14 400.00 |
|  | **TOTAL** | | **R445 420.00** |
|  | **Miscellaneous** | |  |
| 1 | Documentation printing | | R3500.00 |
|  | **TOTAL** | | **R3500.00** |
|  | **GRAND TOTAL** | | **R804 635.24** |

# **INVOICING SCHEDULE**

|  |  |  |
| --- | --- | --- |
| **Phases and Payment** | **Date** | **Amount (R)** |
| **Phase 1:** |  | **R3 339.80** |
| **Phase 2:** |  | **R564.20** |
| **Phase 3:** |  | **R145 364.20** |
| **Phase 4:** |  | **R20 410.20** |
| **Phase 5:** |  | **R564.20** |
| **Phase 6:** |  | **R88 764.20** |
| **Phase 7:** |  | **R564.20** |
| **Phase 8:** |  | **R91 344.10** |
| **Miscellaneous** |  | **R3500.00** |
|  |  |  |
| **Total** |  | **R354 415.10** |
| **VAT 15%** |  | **R53 162.27** |
|  |  |  |
| **Total Contractor tariff** |  | **R445 420.00** |
|  |  |  |
| **GRAND TOTAL** |  | **R852 997.37** |

## 12.1. Terms and Conditions

The terms and conditions are proposed by the subcontractors of NightCore Mech.

The tender response is based on the understanding of NightCore Mech regarding the scope and deliverables of the project. The subcontractors of NightCore Mech work at an hourly rate. The costs are based on deliverables stated in the budget at Section 11: Budget, from page 17. The budget could be subject to change if the clients alter the scope of the project or add additional requirements to the project. The altered budget should be approved by the client. The budget only contains items that the client is responsible for.

## 12.2. Purpose and Context

The client has fully disclosed the contract to NightCore Mech. The client will use the results of the contract work only for the purpose disclosed to NightCore Mech unless otherwise agreed. The client agrees to respect and value NightCore Mech's connection with its suppliers/specialists.

## 12.3. Validity of Submission

The cost estimate is valid for fourteen working days after the submission date.

## 12.4. Contractual Relationship

The parties agree that the Standard Terms and Conditions will govern any contractual agreement between them. When this proposal's approval is contingent on the signature of a Service Level Agreement (SLA), the Standard Terms and Conditions must be considered when creating the SLA. If acceptance of this proposal is contingent on both parties signing a Service Level Agreement (SLA), no work will be undertaken until both parties have signed the SLA.

## 12.5. Duration

The agreement becomes effective after both parties have signed it. The agreement will be valid until the completion date indicated in the scope of the contract work. This is subject to provisions that provide for earlier termination. If the agreement is terminated, the clauses relating to intellectual property and confidentiality will remain in effect.

## 12.6. Client’s Responsibility

Individual full briefings will be provided by client representatives, only by appointment, and access to necessary documents, personnel, and facilities will be permitted.

The customer will provide NightCore Mech with all paperwork and other material (source code) that could be useful in developing the new system.

The client will make their facilities available to NightCore Mech at no cost.

Given the cooperative governance, identified stakeholders should work together. The procedure will be maintained once a chance to engage has been provided and a constructive response has not been obtained promptly.

Comments on draft papers and findings will be delivered to the client within a period that is mutually agreed upon, but not more than seven days.

Draft reports will be given to the client directly, who will receive one set of responses from the bidder. The bidder will include the responses in their final submission and send them to the appropriate authority. Responses have to be received within a week.

## 12.7. Invoicing and payment

NightCore Mech will produce the invoices, and payment will be made directly to them.

Each invoice will be accompanied by proper paperwork for payment.

Within 30 days of receipt of invoices, all accounts will be settled. Interest will be charged at the banking rates set by NightCore Mech.

All sums due under the agreement must be paid by the due date specified in the agreement, or within thirty (30) days after the date of invoice issuance if an invoice is required. Interest will be charged on balances that are past due at prime plus 2%. (2 per cent).

## 12.8. Risk

Intellectual property given to NightCore Mech by or on behalf of the Client according to the agreement will be received, retained, and used at the owner's risk. Unless otherwise specified in the agreement, intellectual property becomes the property of NightCore Mech upon the termination of the agreement.

## 12.9. Confidentiality and publication

All information disclosed to each other during the term of the agreement regarding or arising from the contract work shall be treated as strictly confidential, and the Parties shall not disclose any such information to third parties without each other's written approval, which consent shall not be withheld unreasonably. Consent may not be delayed for longer than two years when the NWU publishes for scholarly purposes. The Parties are not permitted to postpone the submission and examination of theses and dissertations, as well as the conferral of degrees.

## 12.10. Liability

* 1. The Client indemnifies NightCore Mech, and NightCore Mech indemnifies the Client, against any claims that may arise because of the other's negligent act or omission, or non-compliance with any element of the agreement throughout the course and scope of contract work.
  2. Any claim for damages against NightCore Mech based on the agreement is limited to the lesser of the contract price or the amount paid by the Client to NightCore Mech in respect of contract work performed in accordance with the Agreement.

## 12.11. Intellectual property

Unless otherwise indicated in the agreement, intellectual property arising from contract work can be transferred from the Registrar on the terms of an agreement previously signed by the University and NightCore Mech.

## 12.12. Breach

If either party breaches its obligations under the agreement and fails to remedy the breach within seven days of receiving written notice requiring it to do so, the other Party shall have the right, at its option but without prejudice to any other or additional rights or remedies, to cancel the agreement.

## 12.13. Termination

Both parties may mutually choose to dissolve the arrangement. If the client terminates the project for any reason, or by mutual agreement, all direct expenses and professional fees will be billed up to and including the project's termination date.

## 12.13. Waiver

No leniency or exemptions granted to the other party, or a party's failure to exercise its rights as a result of a breach of the agreement, will affect the rights of the first-mentioned party, nor will it be hindered from exercising those rights.

## 12.14. Total Agreement and Amendments

The agreement represents the parties' entire agreement, including any additional restrictions, proposals, or pledges, whether verbal or written. Unless reduced to writing and signed by both parties, no modifications to the agreement will have any validity or effect.

## 12.15. Governing Law

The Republic of South Africa's laws will govern the agreement and will be used to interpret it.

## 12.16. Domicilia Citande et Exencutandi and Notice

For all purposes of and in connection with the agreement, the parties choose the addresses mentioned therein or such other address as may be communicated in writing by one party to the other as their domicilia citandi et executandi. Any notice required by the agreement must be delivered to the applicable party at its domicilium in writing.

## 12.17. Precedence

If these General Conditions and any term or condition in the agreement conflict, the latter will take precedence.

# **RELEVANT EXPERIENCE**

|  |  |  |  |
| --- | --- | --- | --- |
| **Client Name** | **Brief Description** | **Timeframe** | **Members** |
| Janet Liebenberg | Lecturer Leave Day repayment system | 2020: March to April | Hanno Visagie  Michael Rosin  Llewellyn Anthony  Hano Strydom  Annika du Toit  Shené Boshoff |
| Linda Redelinghuis | ATLAS Courier System | 2020: Feb - October | Hanno Visagie  Michael Rosin  Llewellyn Anthony  Hano Strydom  Annika du Toit  Shené Boshoff |
| Rodney Sebopelo | FNAS System/Database | 2021: June - July | Hanno Visagie  Michael Rosin  Llewellyn Anthony  Hano Strydom  Annika du Toit  Shené Boshoff |
| Zander Boonzaaier | Vitality Web App that manages Active Rewards | 2021: August - October | Hanno Visagie  Michael Rosin  Llewellyn Anthony  Hano Strydom  Annika du Toit  Shené Boshoff |
| Zander Boonzaaier | Social Media Web application | 2021: November - November | Hanno Visagie  Michael Rosin  Llewellyn Anthony  Hano Strydom  Annika du Toit  Shené Boshoff |

# **CONTACT DETAILS OF BIDDER**

|  |  |
| --- | --- |
|  | **Zander Janse van Rensburg:**  **Senior Client Representative Agent**  **Office: 018 991332**  **Email: Zander.JanseVanRensburg@nwu.ac.za** |
| A person smiling for the camera  Description automatically generated with medium confidence | **Prof. Neels Kruger:**  **Teaching & Research Focus: IT Management, IT Governance and Knowledge Management at NWU**  **Office: 018 299 2533**  **Email: Neels.Kruger@nwu.ac.za** |

# **ANNEXURE A - CURRICULUM VITAE OF PROJECT MANAGEMENT TEAM**

Llewellyn Anthony:

<https://mega.nz/file/7TZVzBhQ#CbzPHlfV6oLVODK3PrFoFbckQ3tttuKEzYVEeAAFRVk>

Michael Rosin:

<https://drive.google.com/file/d/1_X8QP_H3RFPiL_7yLLWG2OJYNyVxZGPm/view?usp=sharing>

Shené Boshoff:

<https://drive.google.com/file/d/1ciFxdNJw5qj4mSuX9XcG7c0eXrpRto0_/view?usp=sharing>

Hano Strydom

<https://tinyurl.com/HanoStrydomCV>

Annika du Toit

<https://drive.google.com/file/d/1J9Y7Y3yHzVcm6p6LROYVasL_38KFoiet/view?usp=sharing>

Hanno Visagie

<https://drive.google.com/file/d/1456gNARcK6MxcMuv1W0mW8DvSSM6JylD/view?usp=sharing>

# **ANNEXURE B – PROJECT MANAGEMENT**

# **Work Breakdown Structure**

## Phases

## Activities

### **Initiation Phase**

### **Planning Phase**

### **Execution Phase**

### **Project Review/Project Closure Phase**

### **Maintenance Phase**

## Tasks

### **Initiation Phase**

### **Planning Phase**

### **Execution Phase**

### **Project Review/Project Closure Phase**

### **Maintenance Phase**

## Milestones

## Effort

# **Project Plan**

## Schedule

## Dependencies

## Assumptions

## Constraints

# **Project Scope and Goals**

The aim of this project is to design and develop a web-based Python Flask application that will assist the Registrar and lecturers in identifying contract-cheating and original authorship of assignments. The goals of the project are as follow:

* Design and develop a simple yet intuitive interface for users to easily navigate the application, upload source and evidence texts, perform stylometric analysis and text comparisons, generate feedback reports that indicates the level of contract-cheating and authorship
* Increase the accuracy of detected plagiarism levels/authorship attribution.
* Streamline the contract-cheating investigation process of the Registrar/client
* Stylometric functionalities should be implemented within the application. Python statistical libraries should be used to assist those functionalities.
* Stylometric analysis should be based on the following performance metrics:

1. Word, characters and punctuation counts
2. Document metadata
3. Binongo-Analysis and clustering
4. Distance between clusters (Burrow’s Delta)
5. Number of document revisions

* Cross entropy is plausible method to identify the variance in writing styles between authors to identify contract-cheating
* A local corpus should be stored on a device, containing the source text and some additional evidence texts (about ten to eleven documents in total).
* Automation of feedback reports generation. Reports should be generated as soon as possible.
* Automation of identifying text similarities (no manual document comparisons and text highlighting)
* Side-by-side document comparisons in a new window (no manual document comparisons)
* Feedback reports should indicate whether an author’s written piece is an original piece of work, along with the evidence based on the stylometric performance metrics.
* Feedback reports should contain data visualisation methods, such as graphs or charts, to indicate authorship attribution via clustering, and providing the charts with a legend to facilitate comprehension of the charts.
* Use machine learning algorithms (such as n-KK and SMO) to assist the stylometric analysis in order to identify contract-cheating and original authorship of documents.
* The application should be memory-efficient and computing resources should be used sparsely as possible.
* Users should be provided with the option to compare one source document and one evidence text, or one source document and a local corpus on the device.
* The user should be provided with the option to store feedback reports locally on their device or to a cloud service
* Deliver documentation of a high-quality standard that will allow future developers to easily maintain the application

A business case and feasibility study have been conducted to identify the level of viability of the project scope and goals. Should scope changes occur, meetings will be held and documented with the relevant stakeholders. Client satisfaction is part of the Scope Management. The development team will have to re-engineer the previous development team’s application to adhere to user requirements.

# **Deliverables**

## Documentation Oriented Deliverables

## Secondary Documentation Deliverables

## Primary Deliverable

## Secondary Deliverables

### **User Manuals**

### **User Training**

## Stakeholders

# **Critical Success Factors**

The success of a project is determined by several things. The project will be doomed if the team fails to adhere to any one of these factors. The following list attempts to include all essential success criteria.

|  |  |
| --- | --- |
| **Critical success factor** | **Description** |
| Budget | The project budget should be less than the planned budget for the project to be successful. |
| Schedule | The project team should strive to finish the project on time. |
| Stakeholder satisfaction | The end product must satisfy the stakeholder. This can only happen when the project team and the stakeholder interact regularly. |
| Stakeholder time constraints | The project is finished according to the stakeholder's specified timeline. |
| Scope | The project's specified scope should not be modified unless the stakeholder is consulted. |
| Deliverable is on time | The client receives the final version of the deliverable on time. |
| Deliverable quality | The client should be pleased with the project’s quality. |
| Quality standards | The final product's quality is on par with the predetermined quality. |
| Minimum performance specification | The product can run on the client's devices, or it can run with minimal requirements if the client's devices aren't specified. |
| End-user satisfaction | The end-user must use the deliverable and be satisfied with the way it performs. |
| Stakeholder accepts schedule changes | A new schedule must be produced and presented to the stakeholder if any schedule modifications would affect the delivery of the final product. The revised schedule must coincide with the plans of the stakeholders for them to accept it. |
| Stakeholder accepts budget changes | If the budget changes, the stakeholders must be informed, and the team and stakeholders must agree on the new budget. |
| Environment disruptions | To avoid external influence on the project, the working environment must be safe and secure. Precautions should be taken to guarantee that there are no internal interruptions. |
| Major risks do not occur | To avoid big threats from occurring, the environment is adequately safeguarded. |
| Major risks are mitigated | Risks are identified and avoided before they occur. |
| Major risks are well managed | When severe risks do arise, they are carefully addressed and controlled to ensure that they do not jeopardise the project's vital components, such as the team or the devices on which the project is stored. |

# **Risk Plan**

Only a broad range of potential risks will be considered. Because it is impossible to establish a risk strategy for every potential project risk, only the risks that influence the important success elements will be included. The following table assumes the likelihood and impact of the hazards. Because of the nature of the problem, determining the probability and effect of the following risks is extremely challenging, as they may differ for each project. The ratings are on a scale from 1 to 10, with 1 being the lowest and 10 the highest.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Risk description | Probability (1-10) | Impact (1-10) | Priority (1-10) | Triggers | Response plan |
| Overspending | 4 | 10 | 10 | Budget items are more expensive than anticipated. | Follow the budget precisely to avoid unnecessary spending. |
| Behind schedule or schedule changes | 7 | 10 | 10 | The project work takes longer than expected or there are changes to the schedule. | Ask the stakeholder for an extended due date. |
| Stakeholder is unsatisfied | 3 | 8 | 9 | The stakeholder is unsatisfied with the deliverable. | Have regular meetings with the stakeholder to receive feedback. |
| Scope creeping | 5 | 7 | 6 | The project scope is changed or increased. | Any changes to the scope should be approved by every team member and stakeholder before they can be implemented. |
| End-user is not satisfied | 3 | 8 | 5 | The end-user is unsatisfied with the deliverable and refuses to use it. | Provide training sessions and a user manual for the end-user. Take note of any feedback the end-user has to improve on the deliverable. |
| Budget changes | 2 | 10 | 4 | The stakeholder changes the budget. | Working with a modular plan will ensure that the budget plans do not interfere with the work that has already been completed. |
| Safety risks | 1 | 4 | 4 | The working environment is hazardous, and accidents might occur, disrupting the schedule. | Check the workplace for any potential health and safety hazards. |
| Bad quality control | 3 | 9 | 8 | The project quality is not monitored during the development of the project. | Do regular quality tests. |
| Bad relationship with the stakeholder | 4 | 6 | 7 | The stakeholder is unsatisfied with the project team, which could mean they are unsatisfied with the deliverable. | Communicate with the stakeholder frequently and attempt to keep them satisfied. |

# **Project Plan Gantt Chart**

# **Quality Plan**

# **Financial Plan**

## Labour

## Equipment

## Schedule

## Assumptions

## Financial Processes

## Activities

## Roles

## Documents

# **Procurement Plan**

## Procurement Requirements

## Requirements

## Market Research

## Schedule

## Assumptions

## Constraints

## Activities

## Roles

# **Resource Plan**

## Labour

## Equipment

## Assumptions

## Constraints

# **Communication Plan**

## Stakeholder List

## Stakeholder Requirements

## Assumptions

## Constraints

## Activities

## Roles

## Methods, Documents and Technology

# **Appendix C: Nightcore Mech Meeting Minutes**

|  |  |
| --- | --- |
| Program/Area: | SimiLabs Plagiarism and Stylometry Checker |
| Meeting Purpose: | Initiation Meeting |
| Meeting Date: | 2022/03/31 |
| Meeting Time: | 16:00 – 17:00 |
| Meeting Location: | Distributed Virtual Meeting |
| Meeting Facilitator: | Ricus Warmenhoven  Zander Janse van Rensburg |
| Attendees: | Hanno Visagie  Hano Strydom  Michael Rosin  Llewellyn Anthony  Annika du Toit  Shené Boshoff |
| Minutes Issued By: | NightcoreMech |

| **Next Steps:** (Task, Assigned to, Checkpoint Date) | **Owner** | **Due Date** |
| --- | --- | --- |
| Development Team and Stakeholders introduction:  Similabs Plagiarism and Stylometry checker | Development Team  Ricus Warmenhoven  Zander Janse van Rensburg | 31 March 2022 |
| Initiation of Stakeholder Meeting times | Development Team  Ricus Warmenhoven  Zander Janse van Rensburg | 31 March 2022 |
| Initiation of Stakeholder Meeting venues | Ricus Warmenhoven | 31 March 2022 |
| Access to source code | Zander Janse van Rensburg | 31 March 2022 |
| Plagiarism checker resources | Zander Janse van Rensburg | 31 March 2022 |
| Initiation of Project Scope | Development Team  Zander Janse van Rensburg | 31 March 2022 |
| Meeting Minutes Closure | Development Team  Ricus Warmenhoven  Zander Janse van Rensburg | 31 March 2022 |

| **Decisions Made:** (What, Why, Impacts) |
| --- |
| 1. **Initiation of Project Scope:**    * Limit contract-cheating and academic misconduct on a tertiary level    * Design and develop an application that uses stylometry and provide feedback reports that can accurately identify contract-cheating and academic misconduct    * Text-comparisons and similarity will only be a minimal feature of the application (a plagiarism checker)    * Improve the workflow of the application    * Automate the process of identifying academic misconduct, and attempt to limit manual feedback report editing    * Identify what the client’s requirements are (main objectives) 2. **Initiation of Stakeholder Meeting times**    * The attendees decided on meeting Thursdays from 16:00 to 17:00, unless communicated otherwise in advance by the stakeholders. 3. **Initiation of Stakeholder Meeting venues**  * The attendees decided on meeting at the G4-G15 seminar room, NWU, Potchefstroom Campus  1. **Meeting Closure**  * All attendees agreed to the items discussed in the minutes. |

| **Discussion:** (Items/Knowledge Shared) |
| --- |
| **Stakeholder Meeting Venues:**   * Attendees decided on having physical meetings at the G4-G15 seminar room at the NWU, Potchefstroom Campus. Virtual meeting will only be held when communicated in advance by the stakeholders   **Stakeholder Meeting Times:**   * Thursday mornings was ruled out due to some attendees having Honours classes during that time. Ricus Warmenhoven and Prof. Neels Kruger also had other priorities during that time. Thursdays 14:00 to 15:00 was also ruled out, because the development team has ITRI671 class during that time.   **Access to source code:**   * Zander Janse van Rensburg provided login details to the application and Linode server. The development team needs to contact Mentje Gericke from the BMI school to access source code ([Mentje.Gericke@nwu.ac.za](mailto:Mentje.Gericke@nwu.ac.za)). Contact Robert Maxwell ([Robert.Maxwell@nwu.ac.za](mailto:Robert.Maxwell@nwu.ac.za)), the previous project leader of SimiLabs for access to source code.   **Plagiarism Checker resources:**   * Zander recommended the development team to have a look at the CopyLeaks API to understand how plagiarism checkers and the source code behind it works. The development team must submit a sample document and study the plagiarism detection levels. The development team must compare the results from CopyLeaks to Turnitin to determine the gaps between the respective plagiarism checkers, as to decide how a plagiarism checker’s functionalities ca be improved. The development team must have discussions with the lecturers on what the shortcomings are of Turnitin (the current plagiarism checker used by academic personnel at the NWU). Contact individuals/departments that have a lot of experience in forensic analysis and stylometry to successfully design and implement a plagiarism checker. Prof. Tiny du Toit, Prof. Neels Kruger and Mrs. Susan Campher are potential candidates for contributing expert knowledge to this project.   **Project Scope:**   * The main objective of this project is to allow lecturers to detect academic misconduct accurately, and if possible, provide forensic analysis on how students can improve their academic writing style. The application should enhance the process of detecting academic misconduct while also increasing the ROI for the NWU. The stakeholder/client (Zander Janse van Rensburg) experienced some difficulties with the current software as well as Turnitin. The client still has to compare texts manually a huge portion of the time when analyzing texts to determine the level og academic misconduct. Turnitin doesn’t currently allow for authorship attribution functionalities, and the current software’s (SimiLabs) stylometry features also happens to be buggy. Finding large enough repositories (corpora) that allows one to detect academic misconduct accurately is also one of the main issues of this project. Existing plagiarism checkers highlight different similarities, and as a result detects different levels of plagiarism. Zander also mentioned that a huge issue is when one excludes a similarity, one cannot include the similarity within the source text again to check how many times it occurs within the source text. Thus, the document must be re-uploaded to revise those similarities. The client also has issues with the feedback reports generated by the software. The client must copy and paste text to revise similarities as well as needing to open Turnitin and the current software in separate windows. The client also needs to manually highlight text on the feedback reports to identify similarities in the source text. One of the objectives of this project is to build an open-source plagiarism checker that the NWU can invest in, instead of the Registrar wasting funds on foreign/outsourced plagiarism checkers to detect academic misconduct of students. The project also places emphasis on identifying risks/computing power/gaps with plagiarism checkers currently used in the industry. What will the implications be of hosting this application? How much server overhead will be used to put this application into production? What is the memory/computing constraints regarding forensic analysis? More importantly: how can these issues be resolved? The data transfer rates of the current SimiLabs/plagiarism checkers are also a bit problematic. The application should thus aim to provide the analysis in the form of feedback reports, while using minimum overhead and at the fastest rate possible. The long-term goal (not part of this scope) is to provide the NWU with domestic forensic analysis software, and to make the software scalable. The medium-term goal is to allow lecturers to conduct investigations on academic misconduct via using the SimiLabs software. Use CopyLeaks API to test how to build plagiarism checkers. Zander offered to provide some funding for tools or APIs the development team needs to use to design and develop the application, given that it is an absolute necessity for the success of the project. The project manager (Ricus Warmenhoven) suggested that the development team shouldn’t rebuild the application from scratch, as there is a lot of time constraints involved. He also recommended that the development team shouldn’t attempt to design and develop a database to persist application data, as it would complicate the designing process of the application. The functional and technical specifications of the application will be discussed at a later stage with the client. Ricus suggested that the development start to design and build the application (prototypes) as soon as possible to overcome time constraints and development issues. The development team should contemplate whether it is preferable to design a plagiarism checker, which is frequently used at tertiary instantiations, or a forensic tool to assist lecturers at universities to identify academic misconduct. Should the focus of this project be on text comparisons/similarities or stylometry and forensic functionalities? The client was concerned that the development team would need to conduct a lot of research to grasp the functioning and implementation of stylometry. The client suggested that the team should rather focus on designing a small corpus and use last-mentioned to conduct forensic analysis to detect academic misconduct via comparing to texts. Text similarity and comparisons should be regarded as a minimum functionality of the application. The accuracy of the feedback reports is extremely important, as it forms the backbone of the evidence investigators provide as to why they made a claim that a student committed plagiarism. The client wants a web-based application to prevent users from having to install different versions of software. Fixing the workflow of the current application is a top priority. The development team should research stylometry and how to implement it within a software environment. The project must assist in combatting contract-cheating and academic misconduct.   **PARKING LOT**: *[Describe any items that may have been deferred for a later discussion]*   1. Contact Mentje Gericke from the BMI School for project resources 2. Contact Robert Maxwell (former project leader) for access to source code 3. Compiling SOPS documentation for the NWU Registrar |

| **Miscellaneous Items:** |
| --- |
| N/A |

| **Items for next minute:** |
| --- |
| Finding project resources   * Contact Mentje Gericke from the BMI School for project resources ([Mentje.Gericke@nwu.ac.za](mailto:Mentje.Gericke@nwu.ac.za)) * Contact Robert Maxwell ([Robert.Maxwell@nwu.ac.za](mailto:Robert.Maxwell@nwu.ac.za)) for access to source code. * Compiling SOPS documentation for the NWU Registrar |

|  |  |
| --- | --- |
| Program/Area: | SimiLabs Plagiarism and Stylometry Checker |
| Meeting Purpose: | Stylometry Meeting Patrick Juola |
| Meeting Date: | 2022/05/25 |
| Meeting Time: | 14:30 – 16:30 |
| Meeting Location: | G4-G15, Potchefstroom Campus, NWU |
| Meeting Facilitator: | Ricus Warmenhoven  Zander Janse van Rensburg |
| Attendees: | Patrick Juola  Hanno Visagie  Hano Strydom  Michael Rosin  Llewellyn Anthony  Annika du Toit  Shené Boshoff |
| Minutes Issued By: | NightcoreMech |

| **Next Steps:** (Task, Assigned to, Checkpoint Date) | **Owner** | **Due Date** |
| --- | --- | --- |
| Meeting Minutes Initiation | Development Team  Ricus Warmenhoven  Zander Janse van Rensburg | 25 May 2022 |
| Development Team Introduction | Development Team | 25 May 2022 |
| Introduction: Stylometry (knowledge on topic): How Stylometry works | Patrick Juola  Llewellyn Anthony | 25 May 2022 |
| COSC 495 (SPTPC: Authorship Attribution) module discussion and material | Hanno Visagie  Patrick Juola | 25 May 2022 |
| Corpora retrieval discussion (preferable to design a database or use a Corpus API) | Hanno Visagie  Patrick Juola | 25 May 2022 |
| Efficient/Effective APIs for implementing Stylometry in Python | Hanno Visagie  Patrick Juola | 25 May 2022 |
| Cross Entropy | Hanno Visagie  Patrick Juola | 25 May 2022 |
| Stylometric Clustering tools: How to implement and data visualization | Hanno Visagie  Patrick Juola | 25 May 2022 |
| Feedback Reports content (essential information that should be displayed on the reports) | Hanno Visagie  Patrick Juola | 25 May 2022 |
| Patrick Article Discussion (Authorship Attribution Using Stylometry and Machine Learning Techniques) | Hanno Visagie  Patrick Juola | 25 May 2022 |
| Parallel Computing for improving application performance | Patrick Juola  Llewellyn Anthony | 25 May 2022 |
| Document Noise Removal | Hano Strydom  Patrick Juola | 25 May 2022 |
| Meeting Minutes Closure | Development Team  Ricus Warmenhoven  Zander Janse van Rensburg | 25 May 2022 |

| **Decisions Made:** (What, Why, Impacts) |
| --- |
| 1. **Stylometry**    * The application will have to be re-engineered to implement the stylometry functionalities.    * Decrease memory overhead of the current application 2. **Corpora retrieval discussion (DB design or Corpus API)**    * The team will not design a database and instead make use of local corpora 3. **Efficient/Effective API implementation for Stylometry in Python**  * The GUI will be simple and user-friendly * The development team will use NumPy and SciPy for the statistical analysis within the application * The application will be developed in Python and relevant libraries will be chosen to provide stylometry functionalities * Text similarity and comparisons will be a minimal feature of the application  1. **Stylometric Clustering tools: Data Visualization**  * The team will make use of the Binongo-Analysis to visualize stylometry data * The team must decide on performance metrics for identifying contract-cheating  1. **Content on Feedback Reports**  * The feedback on reports must be based onperformance metrics for identifying contract-cheating  1. **Cross entropy**  * The team could potentially make use of cross entropy to identify the variance in writing styles between authors to identify contract-cheating  1. **Machine Learning and Algorithms**  * The team will make use of ML techniques and algorithms such as the n-KK and SMO algorithms to detect contract-cheating. Deep Learning will not be implemented in this application.  1. **Parallel computing for improving software performance**  * The team will refrain from using parallel computing as it can become very costly. The team will attempt to write efficient algorithms that are memory efficient.  1. **Meeting Closure**  * All attendees agreed to the items on the minutes and the outcomes thereof |

| **Discussion:** (Items/Knowledge Shared) |
| --- |
| **Stylometry**   * The development team and Patrick introduced themselves. Patrick gave a quick overview on Stylometry: the inner workings and functions thereof. * Stylometry is the statistical examination of discrepancies in literary styles between authors or literary genres. Stylometry is often used in literary studies, historical studies, social studies, information retrieval and a variety of forensic cases and studies. * Stylometry is applicable to implement in this project, as it is a useful tool that can be utilized to detect contract-cheating and academic misconduct on a tertiary level. It can also be applied in computer code and intrinsic plagiarism detection, which detects plagiarism based on changes in writing style within a document.   **COSC 495 (SPTPC: Authorship Attribution) module discussion and material**   * The development team attempted to inquire whether there are resources or textbooks they can acquire on Patrick’s COSC 495 module, to assist them with designing and developing the forensic application. Patrick mentioned that unfortunately he had no resources that the team could use to assist them with the development process.   **Corpora retrieval discussion (preferable to design a database or use a Corpus API)**   * Patrick suggested that the team should develop the application with tools and programming languages that the team are most comfortable with. There appears to be no best practice when designing and developing this application. The team should adopt a similar outlook when comparing documents to detecting contract-cheating. Keeping the size of the corpus as small and the design as simple as possible would be a preferable principle to keep in mind. Patrick proposed that the stylometry application should be developed in Python, since it is a remarkably simple language to learn as well as having numerous statistical and stylometry libraries. Patrick also mentioned that the NLTK Python library was extremely useful for natural language processing functionalities, as well as providing integrated interfaces to most databases. Patrick recommended that the team should not design a database since most of the databases used in forensic cases tends to be exceedingly small, and thus it would be a waste of time and resources to design and build one.   **Efficient/Effective APIs for implementing Stylometry in Python**   * The team should design a basic forensic application that can assist in the process to detect contract-cheating. The application should automate the detection process without an investigator needing to compare text and documents manually. Text comparisons and similarities should be evident immediately and highlighted. Patrick suggested that the team use <https://jalammar.github.io/illustrated-word2vec/> to detect similarities in texts, via making use of vectors. Words with similar meanings would be grouped into the same vector. Zander asked Patrick whether he thought that it would be possible to develop software that detects contract-cheating via the use of basic stylometry packages. * Patrick responded that it is easy to develop an application that compares two documents for text similarity, however, it becomes more complicated to design and develop an application that uses an enormous corpus to detect contract-cheating. Patrick mentioned that an application that uses ten or eleven documents from a corpus to detect text similarity on a single source document would be designed and developed in the same way that comparing only one source and comparison text to each other would be. * Zander mentioned some critiques against Turnitin and why the current software did not fulfil the user requirements. The current Turnitin stylometry functionality is not open-source yet and Zander plans on eventually making the application scalable and managing larger document corpora to improve the accuracy of detecting contract-cheating. * Patrick once again mentioned that the NLTK Python library was a particularly useful tool to use as the primary natural language processor. * NumPy would be the most preferable Python library for doing statistics * SciPy is another useful Python library for mathematics, science, and engineering * Patrick recommended that the development team can use any frontend/GUI technologies that they are comfortable with * The user interface will be standard but intuitive * Patrick suggested that the team design the application to handle a batch mode (a large number of documents should be sent simultaneously instead of a user having to drag and drop each document when trying to upload files) * Patrick notified the team that digital signatures was not a necessary construct within the application to identify documents.   **Cross Entropy**   * Patrick gave a quick overview on Cross entropy. Cross entropy can be characterized as the difference between two probability distributions for a given random variable or set of events. Lower probability events have more information, higher probability events have less information. An event is more surprising the less likely it is to occur, thus lower probability events are surprising, while highly probable events are unsurprising. Because probable events tend to dominate, a skewed probability distribution has less surprise and therefore a low entropy. Events can also be equally likely to occur, thus balanced distributions are more surprising and, as a result, have higher entropy. Cross entropy can be used to predict the probability that an event is going to occur, in comparison to another probability or outcome. Cross entropy can be used to measure and quantify points where two languages or dialects of a language may vary. Cross entropy can thus also be extremely useful in authorship attribution, as it can indicate the differences between two author’s writing styles via the use of probabilities and calculation of entropy. Therefore, stylometry can be used to detect individual differences as well as differences within a populace (age groups, gender, professions, nationalities, education levels etc.).   **Stylometric Clustering tools: How to implement and data visualization**   * Patrick provided one useful resource the team can use to study stylometry: <http://dh.obdurodon.org/Binongo-Chance.pdf>. One can determine authorship attribution by plotting coordinates on a graph, each author’s coordinates having an assorted colour. The x-axis separates the coordinates of the two authors, and their coordinates are clustered together in groups. The first PC clearly distinguishes between the two authors. The second PC lacks additional authorship information. Essentially, this is what the development team should try to develop: to provide clusters of coordinates on a graph as feedback/analysis which can be used to identify contract-cheating via authorship attribution. Patrick explored the number of occurrences of each of the words used in the texts (from Baum and Thompson respectively) using a computer program. The words with the highest frequency of occurrence were chosen. Pronouns, auxiliary verbs, prepositions, conjunctions, determiners, and degree adverbs are examples of function words used in those texts. These parts of speech serve a grammatical rather than a lexical purpose. The books were assigned equal weight after calculating the average rate of occurrence of each word. In a corpus of over a million words, the words at the top of the frequency spectrum were the least context dependent. The words were grouped into a list and their average rate of occurrence within the texts were displayed in parentheses. Auxiliary verbs were conspicuously absent from the list: they are difficult to manage because one root could take more than one derivational suffix.   Personal pronouns were also absent: even a larger variety were present based on case, number, person, and gender. Another reason for their exclusion is that the characters' sex could skew the outcome.  **Feedback Reports content (essential information that should be displayed on the reports)**   * The development team should measure a set of features via using the Binongo-Analysis. The team can compile a list of key words that appears in a document and the average rate of occurrence of each key word. * Based on theaverage rate of occurrence of each key word, the team can assign colours to each level of rate of occurrence, which can determine the significance of the key word occurring in the document (red can for example be used to indicate a higher level of occurrence within a document). * Patrick suggested that the team can use any type of performance metric or grammar class to detect contract-cheating.   **Patrick Article Discussion (Authorship Attribution Using Stylometry and Machine Learning Techniques)**   * Patrick suggested that the team can use ML to detect variations in writing styles between authors, he did mention that the team might experience some difficulties in detecting contract-cheating if the corpus exists of documents that are not relevant by subject. Patrick placed an emphasis on overall similarity as an individual, instead of detecting similarities by group or populace. Patrick made it noticeably clear that the team had to be aware of user requirements and what the software is supposed to do. Therefore, the user input is extremely important to take note of, as it can influence the accuracy of the feedback reports generated by the software. The functionalities of the program should be clearly defined and adhered to. Patrick also hinted that the feedback reports does not have to necessarily indicate who wrote the source text by name, but whether the person who claimed the source text is theirs IS actually the original author. * Patrick discussed an article with the team, Authorship Attribution Using Stylometry and Machine Learning Techniques, written by Hoshiladevi Ramnial, Shireen Panchoo and Sameerchand Pudaruth. Link to the article: [Stylometry and ML](https://d1wqtxts1xzle7.cloudfront.net/54261431/2016_Authorship_Attribution_using_Stylometry_and_Machine_Learning_Techniques-with-cover-page-v2.pdf?Expires=1653836164&Signature=Chvhv7bFb-HbNU1erQ3qvLbsp8bb~9vMUNP6PAIj3T724hTSljPxE2hPEahB3dqCoGRbVWBI8~WGmYMtcwiE7FAZwDFgUIZJF0phVWaadFNXLXi~PQR4Aqw3-SxRdOc11CEN1K~nGxeA5jok3KfQD8JMxORN~nelf-fBk23ppnRTCzQ76wLf~xNE1CkDEIpOesZicdes5783FDCGV~hVuTvSbgENY9Okn~~ljDiw8Jl0Svrirsc5kOjQP4um47MWaH2jhfzxr1CoXIChEoc7zvg8inYPNlZzDs7wCORdJSa8G-nWcTgEJAWMdunk~F13uTC5t-YhgVz2AbQGzl9Pbw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA) * Patrick referred to Figure 1: System Architecture of the Authorship Attribution System and Plagiarism Detection using Stylometry (Ramnial *et al.*, 2016) * Patrick suggested that the team implement ML techniques within the application by using the Figure as a guideline. Pre-processing, feature extraction, and prediction using machine learning classifiers are the three main stages. Each phase includes a number of procedures. * The development team should communicate with the client (Zander) about the features of the application and what he wants the feedback reports to display. * Zander asked Patrick what his thoughts was regarding extracting the metadata from documents to identify contract cheating. Patrick replied that one can refer to Figure 1 from the article, instead having a Metadata analyzer instead of a text analyzer. One can use this functionality to determine what time zones a document was modified in. One can essentially use metadata from time zones to identify whether an author committed contract-cheating: if an author submitted a document within a specific time zone, but the document was modified in a different time zone, this can be a strong indicator of plagiarism. * Metadata such as software versions and the operating systems of devices a document was edited on can also be clear indicators whether contract-cheating occurred. * Metadata differences can thus be used as a set of indicators (features) to detect plagiarism. * Patrick suggested that the Team use k-NN and SMO algorithms to identify authorship attribution. The algorithms are very lightweight and does not consume a lot of memory. * Useful site to read on how to extract document metadata: <https://docs.groupdocs.com/parser/net/extract-text-from-microsoft-office-word-documents/> * Patrick recommended that the team should refrain from using Deep Learning, as it can become incredibly slow and expensive.   **Parallel Computing for improving application performance**   * Patrick suggested that the team should not make use of parallel computing and the HADOOP framework in order to distribute computing time of the application, since one will need more devices in order to parallel compute, which could have cost constraints. Patrick mentioned that writing a data algorithm that processes the data more efficiently might actually solve the problem of memory overhead better than parallel computing. * Patrick also mentioned that memory overhead issues might actually be fixed if one throw some funds in for accessing more computing power (a thoughtful suggestion). * Patrick suggested that the team use Figure 1 from [Stylometry and ML](https://d1wqtxts1xzle7.cloudfront.net/54261431/2016_Authorship_Attribution_using_Stylometry_and_Machine_Learning_Techniques-with-cover-page-v2.pdf?Expires=1653836164&Signature=Chvhv7bFb-HbNU1erQ3qvLbsp8bb~9vMUNP6PAIj3T724hTSljPxE2hPEahB3dqCoGRbVWBI8~WGmYMtcwiE7FAZwDFgUIZJF0phVWaadFNXLXi~PQR4Aqw3-SxRdOc11CEN1K~nGxeA5jok3KfQD8JMxORN~nelf-fBk23ppnRTCzQ76wLf~xNE1CkDEIpOesZicdes5783FDCGV~hVuTvSbgENY9Okn~~ljDiw8Jl0Svrirsc5kOjQP4um47MWaH2jhfzxr1CoXIChEoc7zvg8inYPNlZzDs7wCORdJSa8G-nWcTgEJAWMdunk~F13uTC5t-YhgVz2AbQGzl9Pbw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA) article as a guideline, because it is a lightweight analysis, one only needs to store fifteen pieces of information per document, thus the memory footprint is tiny   **Document Noise Removal**   * Patrick responded that document noise removal might not actually improve the performance of the application. Patrick recommended to refrain from removing citations from documents, as this is an incredibly useful feature to identify plagiarism. * For instance, if one can detect that sources within a document have been referenced by two different referencing styles such as APA and Harvard, it can be a good indication that plagiarism was commited.   **PARKING LOT**: *[Describe any items that may have been deferred for a later discussion]*   1. Project Demo 2. Future Meetings with Patrick Juola 3. Research document noice |
|  |

| **Miscellaneous Items:** |
| --- |
| * Team needs to improve organizational and communication skills * Communicate Agenda items and meeting questions to client in advance |

| **Items for next minute:** |
| --- |
| Project Demo   * Discussion of the project: GUI and data flows * User requirements from client * Feedback reports   Future Meetings with Patrick Juola   * Patrick informed the team that he would be happy to meet again when the team have made further progress and can provide a physical prototype with some basic stylometry functionalities   Discuss document noice |

|  |  |
| --- | --- |
| Program/Area: | SimiLabs Plagiarism and Stylometry Checker |
| Meeting Purpose: | Project Demo: User Requirements Meeting |
| Meeting Date: | 2022/05/26 |
| Meeting Time: | 13:00 – 15:00 |
| Meeting Location: | E8-G44, Potchefstroom Campus, NWU |
| Meeting Facilitator: | Prof. Neels Kruger  Ricus Warmenhoven  Zander Janse van Rensburg |
| Attendees: | Hanno Visagie  Hano Strydom  Michael Rosin  Llewellyn Anthony  Annika du Toit |
| Minutes Issued By: | NightcoreMech |

| **Next Steps:** (Task, Assigned to, Checkpoint Date) | **Owner** | **Due Date** |
| --- | --- | --- |
| Meeting Minutes Initiation | Prof. Neels Kruger  Development Team  Ricus Warmenhoven  Zander Janse van Rensburg | 26 May 2022 |
| Application GUI | Hanno Visagie | 26 May 2022 |
| User requirements: Zander Janse van Rensburg | Zander Janse van Rensburg | 26 May 2022 |
| Document Metadata | Zander Janse van Rensburg | 26 May 2022 |
| Performance metrics of application | Zander Janse van Rensburg | 26 May 2022 |
| Feedback Reports | Zander Janse van Rensburg | 26 May 2022 |
| Application Data Flows | Hanno Visagie | 26 May 2022 |
| Application Prototype | Development Team | 28 July – 4 August 2022 |
| Meeting Minutes Closure | Prof. Neels Kruger  Development Team  Ricus Warmenhoven  Zander Janse van Rensburg | 26 May 2022 |

| **Decisions Made:** (What, Why, Impacts) |
| --- |
| 1. **Application GUI**    * The client had no issues with the GUI presentation, so the team will stick to the current GUI layout unless the design needs to change. 2. **User requirements and Application functionalities**  * A web-based application must be developed with Python that implements stylometry to assist the process of authorship attribution. * The process of generating feedback reports must be automated, and the statistics/charts/graphs generated by stylometric functions must be indicated on the feedback reports to support evidence of contract-cheating. * Development team must create a corpus that can be compared against the source texts * The client should be able to view documents side-by-side without the need to open documents manually or via opening new windows * The development team must carefully decide on the performance metrics and metadata that is going to be used to contribute to authorship attribution * The team will make use of machine learning algorithms to detect authorship and generate feedback reports. * The team should attempt to split the corpus into smaller parts to improve application performance and memory consumption (1000 words per section)  1. **Document Metadata**    * The development team will make use of document metadata as an additional performance metric to identify the true authors of a document. 2. **Performance metrics**    * Words, Characters, and Punctuation    * Document metadata    * Binongo-Analysis and clustering    * Distance between clusters (Burrows’ Delta)    * Document revisions    * Abovementioned metrics will be used to determine stylometry features 3. **Feedback Reports**    * Feedback report contents will be based on the performance metrics, indicating whether an author submitted an original text. The feedback reports will contain graphs/charts with a legend. The accuracy of the reports is essential, as it forms the backbone of evidence against students accused of plagiarism. 4. **Application Data Flows**    * The team will build the application as indicated in their data flow diagrams presented, the client had no objections. The team will adapt the data flows as necessary 5. **Application Prototype**    * A functional prototype will be delivered to the client between the 28th of July to 4th of August. 6. **Meeting Closure**  * All attendees agreed to the items on the minute and the outcomes thereof |

| **Discussion:** (Items/Knowledge Shared) |
| --- |
| **Application GUI**   * **Login Screen –** a basic login form. Contains textboxes at the center of the screen allowing users to enter their username and password. If the credentials are valid, the user can navigate to the Home screen. The user will be prompted to enter valid credentials until the username or password is correct. * **Home Screen –** contains a menu/navigation bar at the top of the screen, with buttons that allows the user to navigate back to Home, text comparisons, stylometry analysis, feedback reports and finally a logout button. The center of the screen contains a shortcut list of the most recently accessed/uploaded documents. * **Text Comparisons –** contains the menu/navigation bar at the top, with two buttons below that allows the user to either perform a quick or extensive text comparison. Quick comparisons contain a basic file upload form on the center of the screen: a textbox and button for navigating to the source text as well as a textbox and button for navigating to the comparison text. Once the user selected both documents for comparison, they can click on the submit button and the analysis will start. A new window will open that compares the source and comparison document side-by-side.   At the top of the screen there is a navigation bar with buttons that allows the user to navigate back to home or generate a feedback report. Users will have the option to upload the source document to a cloud service or locally onto their device.  Extensive comparisons contain more or less the same layout: a form on the center of the screen, with a textbox and button for navigating to a corpus/corpora (which contains a list of documents that will be used opposed to the source text to identify contract-cheating) as well as a textbox and button for navigating to the source text. Once the user selected both options for comparison, they can click on the submit button and the analysis will start. A new window will open that displays the source document with a form to the right-hand side that lists all the sources from the corpus/corpora that contains text similarities.  Similar to the Quick Analysis, at the top of the screen there is a navigation bar with buttons that allows the user to navigate back to home or generate a feedback report. Users will also have the option to upload the source document to a cloud service or locally onto their device.   * **Stylometry Analysis –** contains a menu/navigation bar at the top of the screen, with a textbox and button for navigating to a corpus/corpora that will be used to identify authorship attribution as well as a textbox and button for navigating to the source text. Once the user selected both options, they can click on the submit button, the analysis will start and display a new screen. The screen contains a result that indicates whether the source document is indeed original property of the author. The screen will also contain forms at the bottom of the screen that indicates the student’s performance, the analysis as to why the author is not the original owner and a chart containing clustered coordinates that supports the evidence from the analysis feedback. The user will have an option to upload the source document to a cloud service or store the document locally on their device. * **Reports –** contains a menu/navigation bar at the top of the screen, with buttons that allows the user to navigate back to Home, text comparisons, stylometry analysis, feedback reports and finally a logout button. The final feedback report will be displayed on this screen.   **User requirements and Application functionalities**   * The process of identifying academic misconduct committed by a student/contract-cheating is as follow: The client receives the alleged plagiarized documentation from the Registrar or SJS, the client uses a plagiarism checker such as Turnitin to check for text similarities between dissimilar sources. Turnitin does apparently possess some stylometric functionalities, these however are not open-source, and data cannot be accessed via third-party applications without the required licensing or permissions from Turnitin. Finally, the client must generate reports based on the text similarities. * The process of generating the reports is currently laborious and involves a lot of copying and pasting from source documents and the compared documents from which Turnitin detected text similarities. The client must manually highlight text similarities in the feedback reports, and document spacing can also become problematic at times. * Currently there does not exist a corpus of documents which can be used to compare against the source documents for text similarities and plagiarism. Thus, the accuracy of the plagiarism levels identified by this process is questionable, and sometimes the client lacks proof of plagiarism since there is currently few to no released software that provides stylometric functionalities (without licensing at appalling prices that is). * The client wants the development team to design and develop an application that implements stylometric functionalities to increase the accuracy of detected plagiarism levels/authorship attribution. The client wants the application to be able to generate feedback reports that indicates the statistics and charts/graphs generated from the stylometric functions to support the evidence of contract-cheating by a party. The client should be able to compare documents side-by-side without the need to open new windows to compare documents. * The generation of the feedback reports must be automated, and the investigation process must become streamlined as a result. In order to identify the authorship of a document, the client suggested that the team should have a look at document metadata, stylometric performance metrics as well as machine learning algorithms, since these key concepts can assist one in identifying contract-cheating. In order to compare documents, the client suggested that it is not necessary to design and create a database, the development team can create a small corpus (local folder on a device) containing ten to eleven documents that can be compared to the source text.   **Document Metadata**   * Metadata is organized reference data that helps to sort and identify attributes of the information it represents. The data included in metadata can be utilized to create timelines, establish explanations, and even more. Metadata, in the hands of a trained digital forensics experts, might give light on a specific issue in a case or perhaps constitute a case's turning point. This is crucial data that can be used to identify contract-cheating. However, metadata on its own, like any other piece of digital evidence, is rarely sufficient to justify a point. Almost all digital evidence necessitates some form of corroboration using a combination of electronic and non-electronic evidence. Without adequately constructing a foundation for that evidence, relying on a single piece of electronic evidence can result in misunderstandings or misinterpretations. Whish is why the development team needs to identify performance metrics of stylometry. Document metadata includes information about the authorship, editing time, and even the computer on which the document was modified, that is stored within a document. Typical metadata of a document includes, but is not limited to: file name, title, author, app name (software document was edited on), version of the software, date created, last printed and saved. Others include the total editing time, person the document was last saved by, word, page and page counts, file path, date the document was created and last edited, date the document was last accessed, file size, document hash etc. One can speculate from these metadata why it can be useful to identify contract-cheating by detecting discrepancies between different document versions. More on document metadata: [MetaData.](../ARTICLES/ADDITIONAL/Chapter-27---Metadata_2012_Digital-Forensics-for-Legal-Professionals.pdf)   **Feedback Reports**   * The client mentioned that the report-generation process is currently time-consuming and includes a lot of copying and pasting from source documents and other documents where Turnitin discovered text similarities. In the feedback reports, the customer must manually mark language similarities, and document spacing can be troublesome to fix at times. The generation of feedback reports must be automated, and as a result, the investigation procedure must be simplified. The client suggested that in order to improve the feedback from the reports, the development team must focus on stylometry performance metrics to support claims of plagiarism from the Registrar.   **Performance metrics**   * Zander visualized the performance metrics of stylometry as follow:  1. **Words, characters, and punctuation**   Zander mentioned that the team can exclude certain words, characters, or punctuations to identify authors of a document. Individual authors tend to use specific grammar or punctuation, by focusing on a specific style/words/punctuation conventions, it might become easier to identify whether an author did indeed write an original text.   1. **Document Metadata**   As mentioned before, document metadata can be extremely useful to identify contract cheating. Last minute to a small number of edits might be a good indication that plagiarism was committed. Different versions of documents can also be a strong indication. If a document was edited on different operating systems by different version of software, it is also rather apparent that a document was sent around and edited by different individuals, rendering an author’s claim of originality invalid.   1. **Clustering (Binongo-Analysis)**   Zander mentioned that by using the Binongo-Analysis, one can use graphs and charts to identify discrepancies between author styles. In the Binongo case-study, the authors were able to decrease fifty dimensions to two using multivariate statistical techniques of a principal component analysis. The best-fitting plane ensures that the most significant information is kept because the first two PCs (principal components) account for the two highest shares of the overall variation. The clusters that form is natural, and the visual representation is accurate. The lower-dimensional representation's interpoint distances give us a picture of the text blocks' relative affinities in the original hyperspace. The authorial component appeared to be the most significant cause of variation in the data sets. Baum's books differed from Thompson's in that the former's are on one side of the first PC's mean, while the latter's are on the other. The Binongo-Analysis confident accurately identified Thompson as the author of the 15th book because of this unwavering constancy.  When PCA is utilized as a graphical tool, it has an advantage over approaches that offer numerical probabilities: PCA's validity as a technique for identifying authors is not based on unverifiable statistical assumptions. Its application has grown in favor among stylometrists, who frequently adopt a multidimensional method in their research. Surprisingly, the approach may be capable of more than only distinguishing between two authors. It can be utilized to isolate differences in literary genre while analyzing the works of a single author. Oscar Wilde's plays were separated from his essays on the best-fitting plane in a prior analysis (Binongo and Smith 1999b). The team can include a legend to identify all items on a chart/graph. More on the Binongo-Analysis case-study: [Binongo](http://dh.obdurodon.org/Binongo-Chance.pdf)   1. **Distance between clusters**   Zander mentioned that the distance between the clusters in a Binongo-Analysis can also be used to identify authors of specific texts. Burrows' Delta and its variants can be used to identify authorship. Feature scaling (several types of standardization and normalization) and the type of distance measures (notably Manhattan, Euclidean, and Cosine) can be used as independent variables and the correct authorship attributions as the dependent variable indicative of the performance of each of the methods. As a result, one can describe in greater depth how each of these two factors interacts with one another and influences outcomes. More on identifying authorship via using Burrows’ Delta: [Understanding and explaining Delta](https://watermark.silverchair.com/fqx023.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAsYwggLCBgkqhkiG9w0BBwagggKzMIICrwIBADCCAqgGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMLZ_BO8Qdp239IrA6AgEQgIICeeG4QgIfFt00DTslStK1g5G2Kac5kPMdYuMsFAOnePJWTZxZemQJ1lB1eHN3fH1DbNmpwaz6MZM1J2JOA2wge4NgYWFL3ux8O4ZE7S6NyOa1eRw1OJPssOP1RGCnGQoJGiEmqYIcGSCNHmcKd3uW8LiSAohuMYje7E-E42PLSZ9V27whDXz3q81GeYpCBKYbaBBPEy9PJzme7E43wwjs6COe2f2i7RLmh1YrXHUd4MZYoB6vB0fwWvMOLuxXb97oaA2dX7GCalPJFj6QwcexF8IQKqIQAoGl3rqiq6wB3NY-Yp-uq4k0oo-exIUTbB6uD6MwW35TiQraX6C6AvNQVMnKJbf8AdN0VTjMduxzTTB_gxDz344HU8-afvDZbtEohPnbqT1AJbJSZChvUPRxMS6LGcKBc6f-m0ItW_cyAO2g-KanwwudiVFSutz4WoF2p10H2dT_r9GIx6uTY7xVCLjDNFZSo1GdiFNpB_Bk1USs4gBThFbOBOBixeLqefnkevt9INx3loXDiZYYRHnycvFgjHoEPSUfDIIEGdosUVbWUy6dONQOWbrdGkwWeIPJ5gqYV5evTHF6jdfQIOimq76_XpxYxh97U4dpBydrMulZCNDkQCxz4IrAGqgeiRWsnODpraSMLNrodcyS5xscWeuP3m3I3xb87y44mFUJDy-IaeQDgETfF5IvGVEFMLYr3FVx5RX-jbMzZkTsGixsnv99Gqo55M7IKSgI3-g_94jKBgdMrGK3WBHLZ5Fdya7cq6pXQdNK3rlING52KD7E1vXh656a2mJU8PXWj9Bxmksdz49P4H30XAxbSLSAoeQvVC_WKclE3LmMXg)  [measures for authorship attribution](https://watermark.silverchair.com/fqx023.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAsYwggLCBgkqhkiG9w0BBwagggKzMIICrwIBADCCAqgGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMLZ_BO8Qdp239IrA6AgEQgIICeeG4QgIfFt00DTslStK1g5G2Kac5kPMdYuMsFAOnePJWTZxZemQJ1lB1eHN3fH1DbNmpwaz6MZM1J2JOA2wge4NgYWFL3ux8O4ZE7S6NyOa1eRw1OJPssOP1RGCnGQoJGiEmqYIcGSCNHmcKd3uW8LiSAohuMYje7E-E42PLSZ9V27whDXz3q81GeYpCBKYbaBBPEy9PJzme7E43wwjs6COe2f2i7RLmh1YrXHUd4MZYoB6vB0fwWvMOLuxXb97oaA2dX7GCalPJFj6QwcexF8IQKqIQAoGl3rqiq6wB3NY-Yp-uq4k0oo-exIUTbB6uD6MwW35TiQraX6C6AvNQVMnKJbf8AdN0VTjMduxzTTB_gxDz344HU8-afvDZbtEohPnbqT1AJbJSZChvUPRxMS6LGcKBc6f-m0ItW_cyAO2g-KanwwudiVFSutz4WoF2p10H2dT_r9GIx6uTY7xVCLjDNFZSo1GdiFNpB_Bk1USs4gBThFbOBOBixeLqefnkevt9INx3loXDiZYYRHnycvFgjHoEPSUfDIIEGdosUVbWUy6dONQOWbrdGkwWeIPJ5gqYV5evTHF6jdfQIOimq76_XpxYxh97U4dpBydrMulZCNDkQCxz4IrAGqgeiRWsnODpraSMLNrodcyS5xscWeuP3m3I3xb87y44mFUJDy-IaeQDgETfF5IvGVEFMLYr3FVx5RX-jbMzZkTsGixsnv99Gqo55M7IKSgI3-g_94jKBgdMrGK3WBHLZ5Fdya7cq6pXQdNK3rlING52KD7E1vXh656a2mJU8PXWj9Bxmksdz49P4H30XAxbSLSAoeQvVC_WKclE3LmMXg)   1. **Document revisions**   The total number of document revisions can also be a strong indicator whether plagiarism was committed. The number of revisions is indirectly proportional to the level of plagiarism.  **Application Data Flows**   * **Login –** user enters username and password. Valid → go to Home page. Invalid → prompt user to re-enter username or password. * **Home –** when user click on the “Home” button on navigation bar, navigate to Home screen**.** * **Quick Text Analysis –** user uploads source and suspected text → software analyses text → generate feedback report → user can save the report locally or on a cloud service * **Extensive Text Analysis –** user selects a corpus → user uploads a source text → source file gets uploaded to the local corpus **→** software analyses text → generate feedback report → report gets uploaded to cloud service * **Stylometry –** user selects a corpus → user uploads a source text → source file gets uploaded to the local corpus **→** software analyses text → generate feedback report → report gets uploaded to cloud service * **Logout –** when user clicks the “Logout” button on the navigation bar, the user is logged out of the application   **Application Prototype**   * The development team agreed that they would build a functional prototype for the client. The prototype will be delivered to the client between the 28th of July to 4th of August. For the time being, the development team needs to research Stylometry, machine learning algorithms and data visualization. The team also needs to sharpen their Python skills and study Flask architecture.   **PARKING LOT**: *[Describe any items that may have been deferred for a later discussion]*  N/A |
|  |

| **Miscellaneous Items:** |
| --- |
| * Prof. Kruger suggested that the team should look at alternative ways to design flow charts. Prof. Kruger mentioned that it should be worthwhile to look at swimline charts, Raci charts as well as IDEF charts, which are useful for visualizing software engineering systems. * The development team should have a look into scattered charts * The development team should read on Art Schneiderman: an expert on process management * Prof. Neels warned the team about the potential dangers of releasing a fully functional prototype: it can lead to minority reports, as predictions/analysis from the software might automatically predict that an individual that is part of a certain group or populace are more likely to commit plagiarism. |

| **Items for next minute:** |
| --- |
| * Discussion of application prototype * Stylometric functionalities * Data visualization techniques * Feedback report contents * Accuracy of analysis * User requirements satisfaction |