**MEETING MINUTES**

**Professional Computing Project Team 03**

**MEETING OVERVIEW**

**Date:** August 11, 2025

**Time:** 12:00PM

**Duration:** Approximately 90 minutes

**Purpose:** Project scoping, task clarification, and workflow planning for document search and retrieval system

**Attendees:**

* Dr. Chris Parsons
* Rania Khan
* Hazel Wang
* James Felstead
* Chunyu Zheng
* Peter Fang

**KEY DISCUSSION POINTS**

**Project Scope Clarification**

The meeting focused on clarifying two main project components and ultimately resulted in a significant scope adjustment. The supervisor explained that the project involves working with approximately **7 million PDF pages** of historical government documents, primarily census data spanning **150 years** across multiple countries.

**Technical Challenges and Constraints**

The team raised concerns about the feasibility of the original two-part task structure:

**Part 1 (Original Scope):** Matching extracted tables to source documents within the 7 million page universe

* **Challenge identified:** Would require OCR processing of the entire document collection
* **Technical constraints:** Computationally intensive and potentially impossible within semester timeframe
* **Team assessment:** Significant technical and time barriers identified

**Part 2 (Refined Focus):** AI-powered web search agent for document location

**Document Security and Confidentiality**

**Critical security requirements established:**

* Materials obtained through official channels while working with the **World Bank**
* **Non-Disclosure Agreement (NDA)** required for all team members before document access
* **Access level:** Read-only access to document universe for reference and development

**Search and Retrieval System Requirements**

**System specifications:**

* **AI-powered web search agent** capable of locating government publications
* **Target sources:** National statistics bureaus, national libraries, university libraries
* **Functionality:** Find documents or identify their physical locations
* **Global scope:** Libraries and archives worldwide

**DECISIONS MADE**

**Task Prioritisation**

**Major scope adjustment:** Team and supervisor agreed to eliminate the complex PDF matching component (Part 1) and focus exclusively on the web search and document location system (Part 2).

**Rationale:** Based on technical feasibility assessments and time constraints

**Access and Security Protocol**

* All team members will sign NDAs before receiving access
* Read-only access to the document universe for reference and search term development
* Secure cloud storage space to be established

**Technology Stack**

**AI Tools Selection:**

* Team will investigate premium AI search tools
* **Funding commitment:** Approximately **$200/month** for enhanced AI research tools
* **Platform preference:** OpenAI's research tools identified as potentially superior to alternatives (Grok, Anthropic)

**User Interface Design**

**Interface approval:** Team's preliminary mockup received strong approval <https://preview--census-explorer-hub.lovable.app>

**Interface specifications:**

* Search fields for country, year, document type, publisher, and other metadata
* Results displaying downloadable PDFs or library location information

**ACTION ITEMS**

**Supervisor Responsibilities**

|  |  |  |
| --- | --- | --- |
| **Task** | **Responsible Party** | **Timeline** |
| Provide NDA documents for team signatures | Dr. Chris Parsons | Immediate |
| Grant read-only access to 7 million page document universe | Dr. Chris Parsons | Upon NDA completion |
| Create secure cloud storage space for team access | Dr. Chris Parsons | This week |
| Provide initial search terms and metadata examples | Dr. Chris Parsons | This week |
| Send details about premium AI search tools | Dr. Chris Parsons | Within days |
| Establish WhatsApp group for efficient communication | Dr. Chris Parsons | This week |
| Provide funding for necessary AI tools and resources | Dr. Chris Parsons | Ongoing |

**Team Responsibilities**

|  |  |  |
| --- | --- | --- |
| **Task** | **Responsible Party** | **Timeline** |
| Research and identify best AI search tools within budget | Student team | Next meeting |
| Sign NDAs upon receipt | All team members | Upon receipt |
| Explore feasibility of JPEG-based document comparison | Peter Fang | Next Meeting |
| Begin preliminary testing with sample documents | Student team | Upon development |
| Develop search functionality using premium AI tools | Student team | Ongoing |
| Refine user interface mockup based on feedback | Student team | Ongoing |
| Test broad search parameters with sample queries | Student team | Upon development |

**NEXT STEPS**

**Immediate Actions (Week 1)**

* **Team familiarization:** Gain comfort with document collection structure
* **Sample searches:** Experiment with preliminary document searches
* **Scope understanding:** Develop familiarity with target document availability
* **Manual testing:** Search for sample documents to understand success rates and challenges

**Weekly Check-ins**

* **Schedule established:** Regular Wednesday meetings
* **Purpose:** Monitor progress, address challenges, iterate on search system
* **Communication protocol:** Immediate contact for blockers rather than waiting for scheduled meetings

**Development Phases**

**Iterative approach:**

1. **Phase 1:** Broad search capabilities development
2. **Phase 2:** Gradual refinement based on results
3. **Phase 3:** Specific search terms and metadata requirements integration

**Expected outcome:** Thousands of documents to be located and categorised over semester

**Long-term Vision**

**Project expansion potential:**

* Current semester focus: Search and location system
* Broader goal: Comprehensive global database of historical demographic data
* **Future opportunities:** Continued collaboration and funding for exceptional work
* **Career prospects:** Possibilities for future employment or expanded project involvement

**PROJECT SPECIFICATIONS SUMMARY**

**Document Universe:**

* **Volume:** 7 million PDF pages
* **Content:** Historical government documents, primarily census data
* **Time span:** 150 years of historical data
* **Geographic scope:** Multiple countries globally

**Technical Requirements:**

* AI-powered search capabilities
* Premium tool integration ($200/month budget)
* User interface with metadata search fields
* Global library and archive search functionality

**Security Framework:**

* NDA compliance required
* Read-only document access
* Secure cloud storage protocols
* World Bank confidentiality standards

**CONCLUSION**

The meeting concluded with mutual confidence in the revised project scope and clear understanding of roles, responsibilities, and immediate next steps. The strategic pivot from technical document matching to intelligent web search represents a more achievable and impactful approach within the semester constraints.

**Meeting Minutes Prepared by:** Rania Khan  
**Date Prepared:** August 11, 2025  
**Next Meeting:** Wednesday (Weekly recurring)