The Australian National University
COMP 6703 eScience Project

Semantic Web for Museums
Project Plan
& Requirements Analysis

February 27th 2006 – June 30th 2006

Version Date: March 22th 2006

Author	Content	Submitted to	Date
Junran Lei U4143677	COMP6703 eScience Project	Client/Technical Supervisor: Tom Worthington Academic Supervisor : Peter Strazdins	22/03/2006

Table of Content

1. Introduction

"The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming. "

-- W3C

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

-- Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001

Additional to the official definition above, one of the most important features semantic web has is the machine-understandable data. Most of the Web's content today is designed for humans to read. Content of Semantic Web is meaningful to computers, which means it can decide what is useful for user automatically.

1.1. Project Context

A week long workshop was presented by Mr Tom Worthington on the use of technology for museums of the Pacific islands region in July 2005. One of the recommendations made following the workshop was to investigate building an on-line repository of materials from across the Pacific. In second semester 2005, Kwok Chung, a computer science student at ANU, undertook a six months project to investigate how this could be done using the Semantic Web. A prototype was produced, which showed that further work on this approach would be worthwhile. The Semantic Web and a digital repository could provide museums a more flexible way to catalog their materials and share digital representations of it.

1.2. Project Content

Although the project is defined as an implementation task, it also involves a few research components. From this point, not only to develop an effective semantic archive system, we also need to publish the research results produced during the development process. The detailed project content would be listed as requirements and possible topics in the appendix. One important issue to the user requirements is that there are no particular users that we can get requirements from. However, we would do some research, which might begin from user interface analysis, on what type of functionalities are most suitable to museum situation.

2. Project Scope

2.1. Objective

Develop an effective semantic web archive system for museums.
Publish the research results produced during the development process.

Time: From 27th February to 30th June

2.2. Deliverable

Project Plan
Software Requirements Analysis
Software Design Description
Test Plan
Code
Installation and Configuration Guide
User Guide
Final Report

2.3. Milestone

Milestone	Date
Initial Presentation & Report	Thursday 03/09/2006
Design Completed	Sunday 04/02/06
Mid-project results due	Thursday 04/20/06
Implementation due	Friday 05/26/06
Final Presentation	Monday 06/05/06
Final Report due	Thursday 06/22/06

2.4. Limits and Exclusions

Implementation would be limited to repositories centralised situation.

2.5. Constraints and Assumptions

Assuming there are no existing systems to be interoperated with in museum.

3. Resource Plan

3.1. Human Resource

The human resource here means the time and effort I can put in the project.

According to the timetable and personal schedule, on average I am able to work 8 hours per day, from Monday to Sunday. However, due to other tasks I am doing at the same time, the effort I can put in the Semantic web project before 04/07/2006 would be just 80 hours approximately. In this case, I have to adjust my schedule to add more workload after this date. Nevertheless, in order to meet the official timetable, I would still produce a prototype before the Mid-project results are due, and leave more development, research and improvement to the second implementation. This adjustment is realisable for the reason that I can dedicate over 300 hours into the Semantic web project after 04/07/2006, while the required workload for the project is 300-350 hours.

We can get more detail from Appendix B Resource Graph. As showed in the graph, the resource is overloaded before 04/07/2006 while underutilised afterwards. So we should move more resource to the period after 04/07/2006.

3.2. Technical Resource

3.2.1. Open source software

1. RDF/Ontology Development Tool: Jena, Protégé

The Resource Description Framework (RDF) is a language for representing information about resources in the World Wide Web. (W3C, RDF Primer)

The OWL Web Ontology Language is intended to be used when the information contained in documents needs to be processed by applications, as opposed to situations where the content only needs to be presented to humans. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. This representation of terms and their interrelationships is called an ontology. (W3C, OWL Web Ontology Language Overview)

Jena (http://jena.sourceforge.net/) is a Java framework for writing Semantic Web applications. It has the features of RDF API, RDF/XML Parser, Reasoning and Ontology subsystem. (Hewlett Packard Laboratories, Bristol, Jena2 Overview)

Protégé (http://protege.stanford.edu/) is a visual development tool for RDF, RDFS and OWL Web Ontology.

The languages and tools listed above are either developing prospectively or widely used in the semantic web technology area.

2. Application and Interface Development Tool: Tomcat, Java

Java language has many excellent features such as unrelated platform, mobility; furthermore, it is compatible with other determining tools. We would develop an application model using Java to communicate interface model, metadata model and repository. As to the user interface development, JSP is compatible with Java very well and most familiar to me, the system programme. So we would use Tomcat and Java to build the interface and application development environment.

3. Repository: Dspace, Fedora

Dspace (http://dspace.org/) and Fedora (http://www.fedora.info/) are the leading repositories, from which we should choose the most suitable and compatible one for our system design and other development environment.

4. Database: PostgreSql, MySql

According to the repository configuration files, PostgreSql and Mysql are the open source databases to be used as backend database of Dspace and Fedora.

5. Others: TBD

We should keep our mind open for seeking other possible development languages or tools and comparing them with the ones on hand to build the better compatible developing environment and then the high quality semantic web system.

3.2.2. Hardware:

TBD

4. Project Schedule

4.1. Work Breakdown Structure

1.0	Project Plan & Report
2.0	Requirement
2.1	Requirement Analysis
3.0	Knowledge Learning and Research
3.1	Development Tools
3.2	RDF
3.3	Ontology
3.4	Repository
3.5	Usable Interface
	(Learning & research throughout all project, but more at the beginning)
4.0	Building Development Environment
5.0	Design Design is actually combined with research and learning. It might include Domain Chart, Object Information Model, State Transition Diagram, Class Communication Model.

6.0	Implementation
7.0	Test
8.0	Project Closeout

4.2. Gantt Chart and Timetable

We use Microsoft Project to produce and manage the project schedule. See Appendix A for the Gantt Chart and Timetable. Please zoom in to read.

As we can see from the Gantt Chart and Timetable, the most important dates of the project are visualised as a diamond shape in the Gantt Chart and the durations of them are listed as zero in the timetable. We can find the same content in Chapter 2.3 Milestone, but it is more concrete and understandable in the project management software. Furthermore, the black bar in the progress bar of tasks represents the actual completion percentage of the tasks, while the red line stands for current date, and the round dot is the planned completion percentage until the current date. We can know about whether the tasks are ahead of schedule or behind schedule from the position of the red line and round dot. For example, the Semantic web project is actually ahead of schedule as the round dot goes beyond the red line.

5. Risk Analysis

5.1. Risk Assessment Form

Risk Event	Likelihood (0-6)	Impact (0-6)
The time required to analyse, design, implement and test underestimated.	2	5
Not enough time to implement all the functions designed.	4	3

5.2. Risk Response Matrix

Risk Event	Contingency plan
The time required to analyse, design, implement and test underestimated.	Reconstruct WBS and reschedule
Not enough time to implement all the functions designed.	Set priorities to the functions and implement them with the priorities order. Leaves the remaining functions to the future plan or future works.

6. Requirements

There are no particular users that we can get requirements from. The requirements

listed below are obtained from the client – Mr Tom Worthington and two reports:
“Report on a Workshop on the Use of Technology for Museums of the Pacific Islands Region 2005” – Tom Worthington
“Digital Heritage for South Pacific Museums Project Findings and Report” -- Kwok Chung, Yew

6.1. Business Requirements

Business Requirements are high level extraction of organization, project and users requirements. It should be used as guide to every activity throughout the project.

1. Promote collaboration and knowledge sharing between museums.
2. Create a resource and knowledge management network or a regional digital archive of materials among museums.
3. Help the Museum Community preserve their distinct cultural heritage knowledge for the future.

6.2.Client Requirements

Client requirements are the most important direction to the general objectives and deliverables of the project, and also the compulsory contents of the project.

1. How to present semantic data (meaningful to computer) to users (meaningful to user). Design a usable semantic web interface specifically for museums.
2. Design or select metadata suitable to museum situation.
3. Design or select RDF, Ontology, repository, and other components suitable to museum situation.
4. What tools and standards should be used? Select suitable tools and standards for the project context.
5. Produce an exercisable and effective online semantic archive system for museums.

For the time limitation of the project, it is more realistic to use exist metadata, RDF, Ontology standards and probably add customised contents to them. As for user interface, we could analysis the existing web interface of museums around the world and other semantic webs, and also apply some theories from user interface development area to our design. According to the analysis in 3.2 Technical Resource, we would develop application model using Java to communicate interface model, metadata model and repository; and implement user interface with JSP on the Tomcat platform. For further technical detail such as candidate languages and tools, design and implementation issues, please refer to 3.2 Technical Resource and future design reports.

6.3. User Requirements

There are just a few user requirements that we can get from the reports. However,

they can still give useful direction and reference to system development.

1. Provide an online digital archive system.
2. For the limited technical staff, equipment and telecommunication links of the region.
3. Access and use by scholarly, non-expert users and general public.

6.4. Possible Research Topic

1. Compare and choose repositories from Dspace, Fedora or other Repositories.
2. How to communicate and transfer data between decentralised organisations. (Webservice? Harvesting?)
3. How to store RDF Model, Document or DB?
4. How to interoperate with existing systems?
5. Can we incorporate existing technologies and components together to produce a usable system?

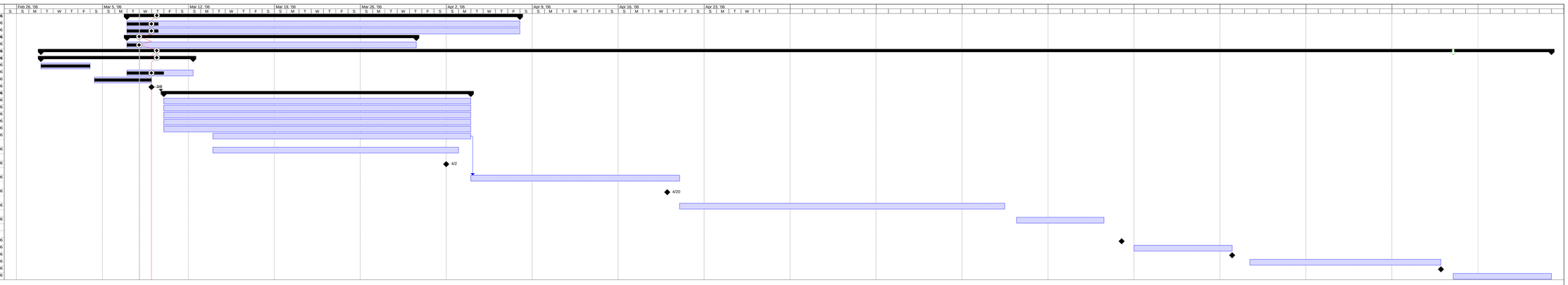
We would choose one to two topics from the above or other topics that come up during the development process. We would not fix the research topics as we do to the implementation topics, because the design and implementing process are actually research process as well since there is so much uncertainty and possibility in semantic web technology area. In fact, designing and implementing is getting our research results and turning it to implementation. Using this strategy could add more flexibility and imagination to the project and also make it more adapted to the time limitation.

6.5 Requirements Priorities

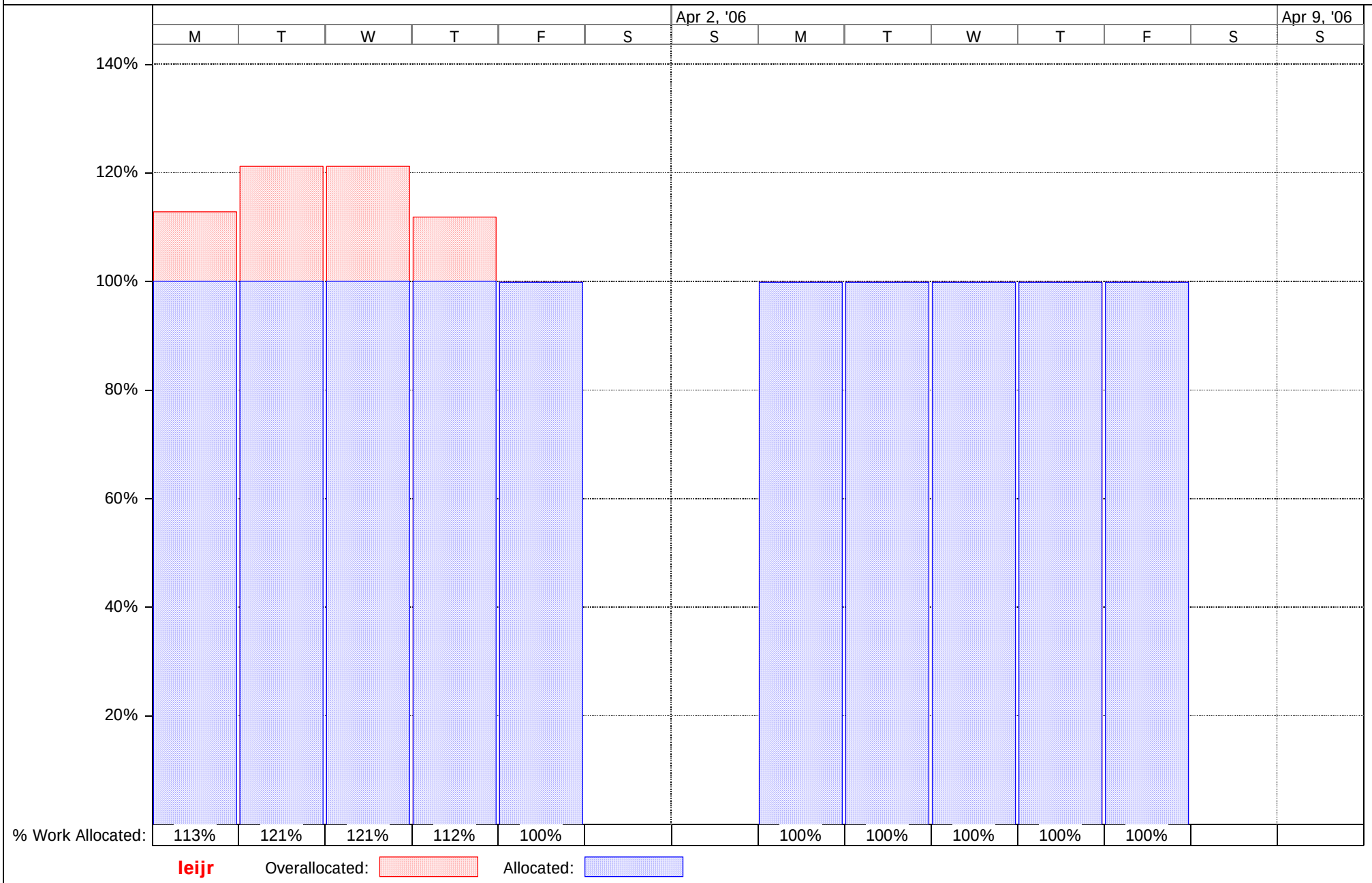
The requirements above are listed with priorities order. All requirements are originated from business requirements. Client requirements are the most important and compulsory components. In order to meet the client requirements at the maximum level, user requirements and research topics would be deducted or cancelled according to time limitation. With the client requirements, producing an exercisable and effective online semantic archive system for museums, the implementation of which must been guaranteed, has the highest priority

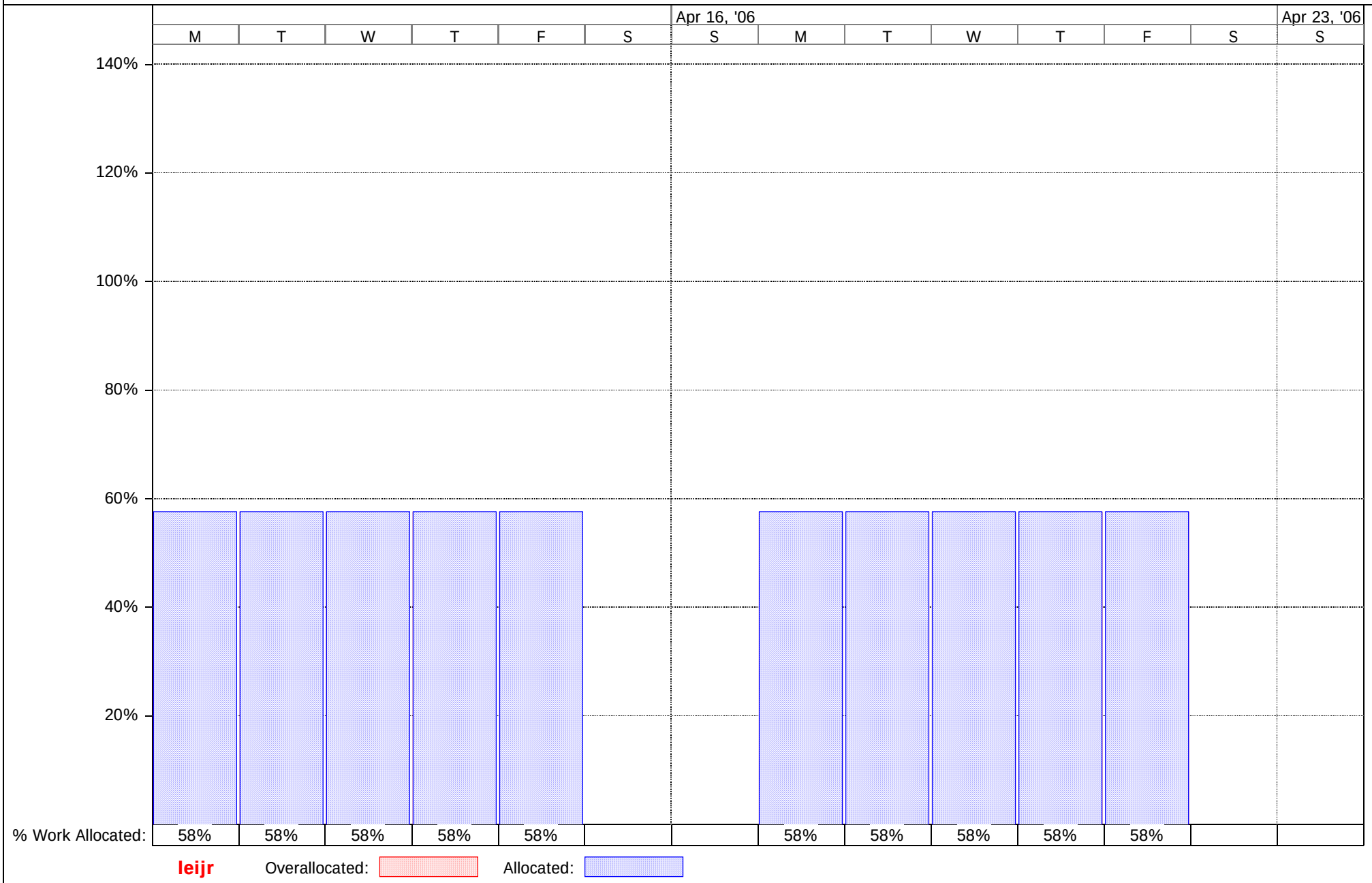
Appendix A Gantt Chart and Timetable

ID	Task Name	Duration	Start	Finish
1	**Software Assignment**	**33.44 days**	**Mon 3/6/06**	**Fri 4/7/06**
2	Software methodology	33.44 days	Mon 3/6/06	Fri 4/7/06
3	Software Architecture	33.44 days	Mon 3/6/06	Fri 4/7/06
4	**Network Course**	**24.94 days**	**Mon 3/6/06**	**Thu 3/30/06**
5	Essay	24.94 days	Mon 3/6/06	Thu 3/30/06
6	**Semantic Project**	**126.88 days**	**Mon 2/27/06**	**Fri 6/30/06**
7	**Requirement**	**14 days**	**Mon 2/27/06**	**Sun 3/12/06**
8	Requirement Analysis	4.69 days	Mon 2/27/06	Fri 3/3/06
9	Requirment Specification	6.81 days	Mon 3/6/06	Sun 3/12/06
10	Project Plan & Report	4.5 days	Sat 3/4/06	Wed 3/8/06
11	Initial Presentation & Report	0 days	Thu 3/9/06	Thu 3/9/06
12	**Knowledge Learning and Research**	**24.94 days**	**Fri 3/10/06**	**Mon 4/3/06**
13	Development Tools	24.94 days	Fri 3/10/06	Mon 4/3/06
14	RDF	24.94 days	Fri 3/10/06	Mon 4/3/06
15	Ontology	24.94 days	Fri 3/10/06	Mon 4/3/06
16	Repository	24.94 days	Fri 3/10/06	Mon 4/3/06
17	Usable Interface	24.94 days	Fri 3/10/06	Mon 4/3/06
18	Building Development Enviment	21.56 days	Mon 3/13/06	Mon 4/3/06
19				
20	Design	21.19 days	Mon 3/13/06	Sun 4/2/06
21				
22	Design Completed	0 days	Sun 4/2/06	Sun 4/2/06
23				
24	First Version Implemetation	17.75 days	Mon 4/3/06	Thu 4/20/06
25				
26	Mid-project results due	0 days	Thu 4/20/06	Thu 4/20/06
27				
28	Second Version Implementation	27 days	Fri 4/21/06	Wed 5/17/06
29				
30	Test & Debug	7.56 days	Thu 5/18/06	Thu 5/25/06
31				
32				
33	Implementation Due	0 days	Fri 5/26/06	Fri 5/26/06
34	Document Organization	7.38 days	Sun 5/28/06	Sun 6/4/06
35	Final Presentation	0 days	Mon 6/5/06	Mon 6/5/06
36	Final Report	16.5 days	Tue 6/6/06	Wed 6/21/06
37	Final Report Due	0 days	Thu 6/22/06	Thu 6/22/06
38	Project Closeout	8.06 days	Fri 6/23/06	Fri 6/30/06



Appendix B Resource Graph


Apr 2, '06
Apr 9, '06
M T W T F S S M T W T F S S
140%
120%
100%
80%
60%
40%
20%
% Work Allocated:
113% 121% 121% 112% 100% 100% 100% 100% 100% 100%
leijr
Overallocated:
Allocated:


Apr 16, '06
Apr 23, '06
M T W T F S S M T W T F S S
140%
120%
100%
80%
60%
40%
20%
% Work Allocated:
58% 58% 58% 58% 58% 58% 58% 58% 58% 58% 58%
leijr
Overallocated:
Allocated:

Reference

1. “Report on a Workshop on the Use of Technology for Museums of the Pacific Islands Region 2005” – Tom Worthington
2. “Digital Heritage for South Pacific Museums Project Findings and Report” -- Kwok Chung, Yew
3. The Semantic Web, Scientific American, May 2001 -- Tim Berners-Lee, James Hendler, Ora Lassila
4. Semantic Web -- W3C
5. Sustainable Online Technology for Museums of the Pacific Islands Region – Tom Worthington
6. http://www.tomw.net.au/2005/emuseums - Tom Worthington
7. http://www.tomw.net.au/2005/dm - Tom Worthington
8. Semantic Web Road map - Tim Berners-Lee
9. Semantic Web demo - http://claimaker.open.ac.uk/Sandpit/
10. Chris Blackall's Research Notepad: http://ibid.typepad.com/phd/technologies/index.html - Chris Blackall
11. Making a Semantic Web - Joshua Allen