

IPRO 327 – Semantic Web News Engine
Project Plan

I. Team Information

A. Team Member Roster

| <u>Member Name</u> | <u>Background</u> |
|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| Daniel Copeland cataphract40@gmail.com | 4 th Year Political Science major, first IPRO |
| Evan Estola estola@gmail.com | 4 th Year CS major, 2 nd IPRO, 2 nd semester in IPRO 327 |
| Max Kaim mkaim@iit.edu | 3 rd Year CS major, first IPRO |
| Jay Mundrawala jmundraw@iit.edu | 2 nd Year CPE major, 2 nd IPRO, 2 nd semester in IPRO 327 |
| Steven Peterson speters9@iit.edu | 4 th year ITM major, 2 nd IPRO (Solar Decathlon) |
| Daniel Price Jr dprice6@iit.edu | 4 th Year CS/CPE major, 2 nd IPRO |
| Ori Rawlings orawling@iit.edu | 3 rd Year CS major, first IPRO |
| Laura Rodriguez lrodrig7@iit.edu | 4 th Year CPE major, 2 nd IPRO |
| Daniel Sirotzke dsirotzk@iit.edu | 3 rd Year CPE major, 2 nd IPRO |

B. Team Member Strengths, Needs, & Expectations:

| <u>Member Name</u> | <u>Strengths</u> | <u>Needs</u> | <u>Expectations</u> |
|--------------------|----------------------------------------------------------------------------|--------------------------------------|---------------------------------------------------------------------------------------------------------|
| Daniel Copeland | Soft skills, teamwork | Aid in technical aspects | Functional semantic search; accomplish all our goals, win FIRST PLACE! |
| Evan Estola | Previous IPRO, semantic web technologies, IR, writing, presenting | Time management | Come together as team, create something all are proud of |
| Max Kaim | Soft skills, critical evaluation, Java | Improve programming/ design | Gain useful understanding of semantic systems. Improve overall coding and designing skills. |
| Jay Mundrawala | Data Mining, IR, System Programming | Time Management, Communication | Have a search engine that has more features than last semesters. |
| Steven Peterson | Focusing, design | Technical skills, | Learning how to work |

| | | | |
|-----------------|---------------------------------------|-------------------------------------|------------------------------------------------------------------------------------|
| | | time management | within a team. |
| Daniel Price Jr | CS, previous IPRO experience | Server skills, time management | Walk away having learned something about search engines and servers |
| Ori Rawlings | IR, semantic web, soft skills, Java, | Improve coding, time management, | Success – set deadlines & meet them, ensure team communicates, learn from mistakes |
| Laura Rodriguez | CS skills, Java | Learn XML, CSS | Know how to design in XML, CSS, improve presentation skills |
| Daniel Sirotzke | HTML, CSS, Javascript, website design | Time management, server side skills | Learn more about servers |

C. Team Identity



Our IPRO group has come to the consensus that we'd like to keep the previous logo from last semester: it consists of a stylized pink oval, made to look like a pig's snout, as a reference to the name 'Swyne' (Semantic Web News Engine). Our team's mascot is a pig who declares that he is 'munching all your news articles', as the search engine currently aggregates news articles.

II. Team Purpose and Objectives

A. Team Purpose

The Swyne Project's objective is to extend the capabilities of the semantic search engine developed in the previous semester of this particular IPRO project. Proposed extensions include the ability to search for "entities" (e.g., people) and their relationships to other entities in a fact database and implementing user interface support for the extensions.

Another goal of the Swyne Project is to work on our collective soft skills. These include, but are not limited to, interpersonal communications, presentations, management functions, and time management. These skills are all critical to not only a successful project, but also a successful career outside of college.

Our search engine will hopefully be an inspiration to others who are interested in working on a semantic search engine in the future and expanding upon our results.

B. Objectives

Our team has three primary categories of objectives. We broke them up into these groups in the hopes that it would more clearly define what we want to achieve, thereby improving our ability to complete them. They are:

- i. Technical Objectives
 1. Entity Search functionality for Swyne Project
 2. Faster, more efficient back-end server architecture
 3. Better organized, more user-friendly interface

- ii. Broader Impact Objectives
 1. Better accountability for our continuity
 2. Stricter use of document sharing & record keeping
 3. More clearly defined tracking of progress

- iii. Soft Skill Objectives
 1. Communication
 2. Interpersonal Relations
 3. Effective Presentations
 4. Management of Personnel
 5. Time Management

III. Background

- A. Our project has no customer or sponsor. Our goal in this project is not to provide a search engine for any specific corporation or company, but instead to create a piece of software that everyone can use and benefit from. In our best case scenario, our search engine would become popular and be used by thousands, if not millions, of users each day. The search engine would provide better results that are more tailored to the user's desires.
- B. However, the project does face several problems before we can successfully release a user-friendly program. The functionality of the search engine is questionable: currently it can only search through archived (or 'canned') news articles that we provide it. The user interface also needs improvement: currently the search engine is somewhat difficult to navigate and requires an overhaul to ensure a more streamlined interface.
- C. Search engines today look at keywords that the user has chosen and compare them to documents that the search engine has gathered over time. The engine, however, has very limited ability to discern the deeper meaning of what the user is searching for. Subtle variations in meaning will be lost to it. Our project will attempt to overcome these problems by using semantic web technology to help ensure quality results while web searching.
- D. Semantic web technology has, so far, seen little implementation in the modern world. Few semantic search engines exist, and those that do are not well known or used often.
- E. Once ethical issue that arises is hosting news articles that we crawl on our website. These articles are copyrighted and may have restrictions on how they may be used. Since this is currently a research project, we do not see any problems. If this were to become a commercial news search engine, we would need purchase the rights to host the news articles.
- F. No business or societal costs are likely to be incurred outside of our own equipment costs, which should be very small (<\$150).

IV. Team Values Statement

- A. Members will be expected to:
 - i. show up for class sessions
 - ii. anticipate deadlines
 - iii. complete assigned portions of task on time
 - iv. communicate freely and often with other team members
 - v. contact team in advance if deadline cannot be met
 - vi. provide input into any/all aspects of project whenever beneficial
 - vii. conduct themselves in a proactive manner, anticipating problems and taking preventative measures to ensure they do not affect the project

Although we are broken up into three ‘sub-teams’, we are still an IPRO group and will work collaboratively towards our main goal. This includes cross-oversight and cross-communication to ensure that all three parts of the projects work together as parts of the same machine.

- B. For the purposes of streamlined and easy communication, all communication will be through our private Swyne wiki and our private IRC channel. We make use of a chain of command:
 - i. Prof. Wai Gen Yee, the faculty leader and advisor of the project
 - ii. an overall team leader (student)
 - iii. three sub-teams each with a sub-team leader
 - iv. the members of each sub-team with their individual responsibilities

V. Methodology/Brainstorm/Work Breakdown Structure

- A. The Problem (in other words, the goal that our team is working towards):
Work as a team to extend the capabilities of an existing semantic search engine.
- B. Our Approach: We will split the Swyne Project members into 3 teams:
 - i. Crawler sub-team ‘listens’ to various news web feeds. As articles are published on the internet, they are downloaded, and the article body is extracted out of the web page. The article is then handed off to the Server Indexer.
 - ii. Server sub-team extracts entities out of news article text and stores data in database. Provide an interface to the stored data for the usability team.
 - iii. Usability sub-team create front end for accessing stored news data. Ensures user-friendliness of the website and online interface, by performing and analyzing end user testing.

Each team will have a team leader, who will report up to the overall group leader. That individual will confer with Prof. Wai Gen Yee to ensure the project remains on track.

- C. Testing will be conducted by a combination of automated testing to ensure proper function of the server and crawler software and user testing will be used to develop iterative feedback for the interface design. Documentation of progress/results will be catalogued and stored on each sub-team’s page on main project wiki.
- D. The Swyne Project will be doing research and testing throughout the course of the semester while we code. Each member is required to do two presentations over the course of the semester on topics relating to the field.

- E. There won't be much in the way of test results – we are coding a program. A functional search engine will be a successful 'result'. The usability subgroup is planning to do human trials, so they will likely perform surveys to ensure user test subjects find the interface acceptable.
- F. IPRO deliverables will be produced by a collaborative effort between team leaders and a designated deliverables producer.

VI. Expected Results

Activities during the project will mainly be coding and software development, documentation and IPRO deliverable generation, and conducting user testing. We expect user testing on our interface prototypes to shed light on what elements are intuitive and easy to use, and which ones are poorly designed. A potential product resulting from the Swyne Project is a web application allowing people to easily search through news articles via a variety of criteria. The group will produce, by the end of the semester, a working semantic search engine capable of returning valid, useable, and most importantly relevant results based off of the user's input. The main challenge that needs overcoming in this project is the meeting of all assigned deadlines, and the successful completion of a viable semantic search engine. Iterative user testing on interface prototypes will give us feedback every two weeks as to what works in the interface and what needs to change. As each prototype is released every two weeks, it will hopefully steadily increase in quality until we reach our final prototype at the end of the semester.

VII. Project Budget

The only anticipated expenses would come from the Server subteam, which would amount to either one or two additional hard drives. Equipment suitable for our needs would come to \$130 + \$5-\$10 for shipping. It is readily available from Newegg:

<http://www.newegg.com/Product/Product.aspx?Item=N82E16822148337>

VIII. Schedule of Tasks and Milestone Events

| <u>Team</u> | <u>Task</u> | <u>Due Date</u> |
|-----------------------|-----------------------------------------------------------|-----------------|
| Crawler Team | Web Feed List | 18 Feb |
| | Text Extractor | 23 Mar |
| Server Team | Define API, Implement Mock Objects | 18 Feb |
| | Entity Search, Ranking, Geonaming, RDF Browser Prototypes | 4 Mar |
| | Integration | 11 Mar |
| | Debugging | 1 Apr |
| | Testing Completed | 15 Apr |
| Usability Team | Research | First 2 Weeks |
| | Prototyping, User Testing | Ongoing |
| | First Prototype | 4 Mar |
| | New Iteration of Prototype | Every 2 Weeks |
| | 3 rd Iteration | 13 Apr |
| I PRO 327 | Midterm Report | 24 Feb |
| I PRO 327 | Final Report | 27 Apr |

IX. Individual Team Member Assignments

A. Crawler Team (create viable crawler program)

- i. Ori Rawlings
- ii. Dan Copeland

B. Server Team (handle server back end, database, functionality)

- i. Evan Estola
- ii. Jay Mundrawala
- iii. Dan Price
- iv. Dan Sirotzke

C. Usability Team (end user testing, UI, front end)

- i. Laura Rodriguez
- ii. Max Kaim
- iii. Steven Peterson

X. Designation of Roles

- A. Minute Taker: Jay Mundrawala/Dan Copeland
- B. Agenda Maker: Prof. Wai Gen Yee
- C. Time Keeper: Prof. Wai Gen Yee
- D. Weekly timesheet collector/summarizer: Prof. Wai Gen Yee
- E. Master schedule maker: Handled via Swyne wiki
- F. iGroups: Not used – our team instead elected to make use of a private wiki that only our team members can access. We handle all communication and progress by updating the wiki. The wiki is located at <http://ipro327spr09.pbwiki.com/>. It requires a username/password to log in, so there is no chance of an outside person vandalizing our progress.