Objective

Our goal is to design a search system that understands the meaning of its indexed data. This will allow users to search for data based on time and place among other semantics.

Approach

We decided to split our group into 3 subgroups – Server, Crawler, and Usability. The Crawler team would focus on finding semantic info and news articles from around the country for the search engine to analyze and archive. The Server team's job was to look at the article text and extract the entities and store them in the database. Finally, the Usability team would ensure that the search engine is as user-friendly as possible.

Team Members

CRAWLER TEAM
Ori Rawlings*   (CS, 3rd year)
Dan Copeland   (PS, 4th year)

SERVER TEAM
Evan Estola**   (CS, 4th year)
Jay Mundrawala*   (CPE, 2nd year)
Daniel Price   (CS/CPE, 4th year)
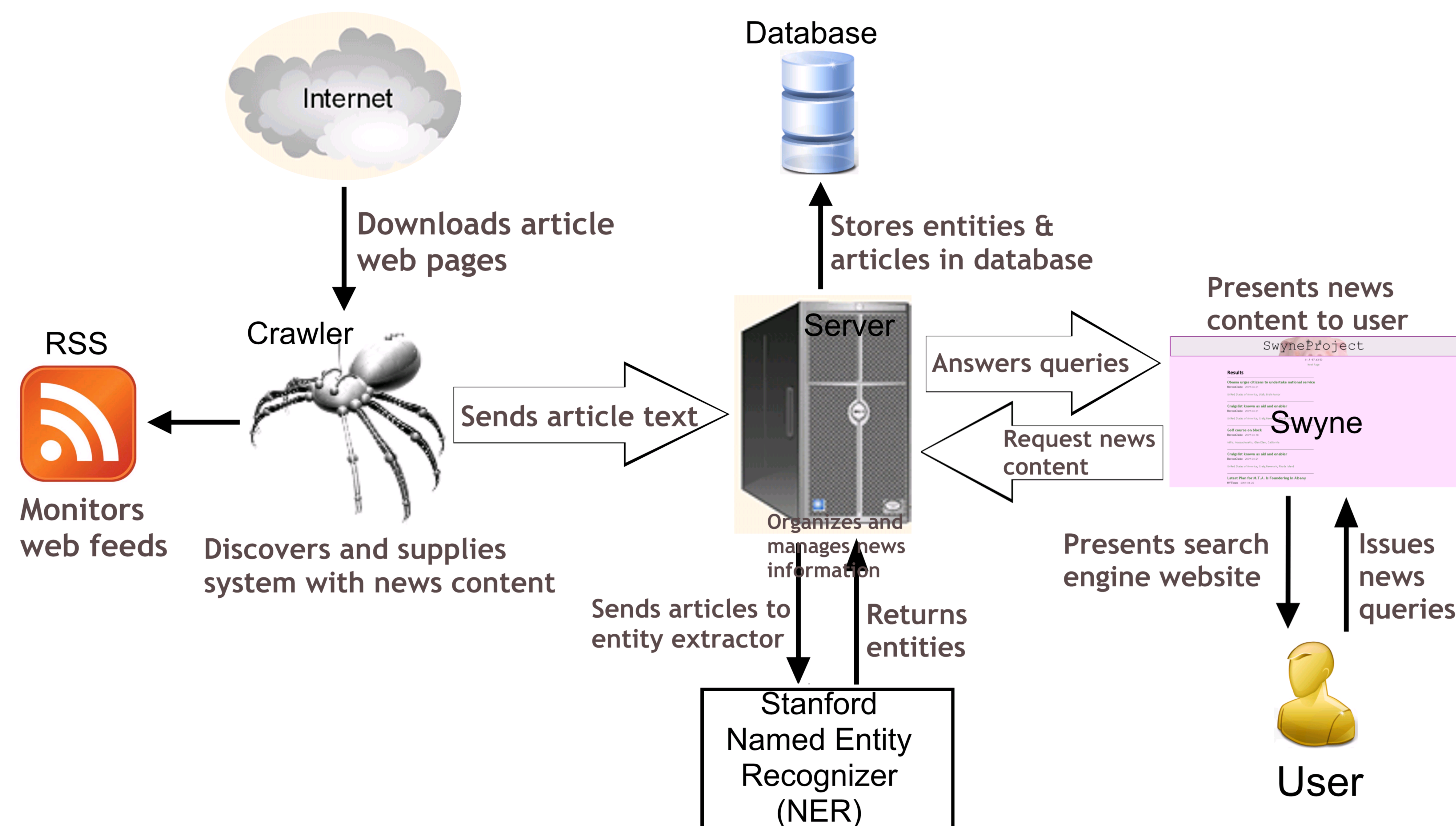Daniel Sirotzke   (CS, 3rd year)

USABILITY TEAM
Max Kaim   (PSYC/CS, 3rd year)
Steven Peterson   (ITM, 4th year)
Laura Rodriguez*   (CPE, 4th year)

*Team Leaders | **IPRO Leader

SWYNEPROJECT

IPRO 327 A Geographically Searchable News Aggregator
http://swyne.homelinux.org
Spring 2009

HOW IT WORKS

Internet
Downloads article web pages
Database
Stores entities & articles in database
RSS
Monitors web feeds
Crawler
Sends article text
Discovers and supplies system with news content
Server
Organizes and manages news information
Sends articles to entity extractor
Returns entities
Stanford Named Entity Recognizer (NER)
Answers queries
Request news content
Presents news content to user
Swyne
Presents search engine website
Issues news queries
User

Group Goals

- Gain valuable experience and leadership skills
- Practice communicating re search results
- Develop open-ended problem solving skills

System Goals

CRAWLER TEAM
- Capability to crawl multiple news sources
- Set up feeds from newspapers across the country

SERVER TEAM
- Enhance identification of entities in news articles
- Improve browsing of relationships between entities

USABILITY TEAM
- Unify project look and feel
- Improve website user interaction

Future Work

CRAWLER TEAM
- Crawler that adapts to new sources

SERVER TEAM
- Optimization for speed

USABILITY TEAM
- Expand user testing