|  |  |
| --- | --- |
| Pegasus%20-%20Black%20on%20White | Department of Electrical Engineering & Computer Science, CS DivisionCollege of Engineering & Computer ScienceUniversity of Central Florida |
| *Return Form to*: Dr. Mark Heinrich, heinrich@cs.ucf.edu |
| **COP 4934: Computer Science Senior Design** |
| **Proposed Project Description Form\*** |
| (Sponsors who are willing and able are asked to provide a Team Donation of $1500 or more for supplies and the running of the CS Senior Design Program) |
| Will support: X Cannot support: |
|   |
| Sponsoring Organization: | Siemens Energy, Inc. |
|  |
| Mailing Address: | 4400 N. Alafaya Trail Orlando, FL 32826 |
|  |
| Project Contact: | Matthew Evans | Position: | Director, Smarter Service |
|  |
| Contact Phone: | 407-736-2055 | Fax: | 407-736-5838 |
|  |
| Contact E-mail: | matthew\_evans@siemens.com |
|  |
| Project Title *(working)*: |  |

*Please feel free to use as much space as needed for each of the sections below.*

|  |
| --- |
| **Background of Company/Organization**(Provide a brief overview of the company/organization and the specific project location here) |
| Siemens Energy Service provides after-sales support to Siemens’ global fleet of gas, steam, and wind turbines. The division’s global headquarters is in Orlando, Florida. |
| **Statement and Scope of the Problem(s)**(Provide the problem statement here; Please be as specific as possible to help us understand the scope of the problem, and thus the scope of the project, specifically the design content) |
| Siemens Energy Service collects continuous sensor and diagnostic data from over 8000 machines. Analyzing this machine data is critical to preventing unscheduled downtime, maximizing availability and reliability, and optimizing performance for customers. Delivering the next level of operational flexibility and performance requires combining machine data with diverse sources of unstructured content that capture historical events and provide critical context to sensor signals. Examples of these unstructured or semi-structured documents include field service reports and repair records.The proposed project tackles a key element of this overall challenge – designing and prototype a pipeline architecture that can:1. Ingesting semi-structured text (such as service reports or repair records),
2. Extracting items and entities (e.g., customer, location) based on a provided ontology,
3. Categorizing issues based on text analysis (e.g., crack in turbine blade), and
4. Populating a relational database with the extracted information.

Building a relational database containing relevant metadata and issues from the documents will enable a host of new applications and make it easier to provide historical information to technicians and customers in the field. |
| **Overall Project Goal(s)**(Describe the overall goals of the project in this space) |
| The overall project goal is to design and implement an architecture that will:1. Store the provided documents (can be provided in Word, PDF, or ASCII text),
2. Load and interpret the provided ontology to understand terms and hierarchies of interest (e.g., “turbines” have “row 1 blades” which can have “coating erosion”),
3. Provide a search function that will return sentences/paragraphs containing words in the ontology,
4. Enable the use of packaged Natural Language Processing (NLP) algorithms to extract related pairs of terms and classify issues by severity (based on a provided training set); and
5. Store results of processes #3 and #4 in a relational database.

Item #4 is considered a stretch goal for the design project. |
| **Project Objectives**(Enter the project objectives that will help achieve the goals of the project; Please be as specific as possible)1. Select a scalable storage platform and search engine for unstructured and semi-structured documents.
2. Select a suitable relational database to store the processed results from the documents.
3. Parse and load the provided ontology (can be provided as a text or XML file).
4. Load and index the provided documents.
5. Demonstrate the ability to search the documents for terms in the ontology.
6. Store results from searches in the relational database (e.g., store all “row 1 blade” issues from the provided reports).
7. Implement at least one packaged NLP routine to process either the raw documents or results from the document search.

Objective #7 is considered a stretch goal for the design project. |
|  |
| **Expected Project Deliverables**(Enter the expected project deliverables including, if possible, proposed project tasks; Please be as specific as possible) |
| 1. Architecture diagram outlining major components (document store, search engine, relational database, etc.)
2. Performance results (search times, scaling tests, etc.)
3. Search results including number of occurrences of each ontology term.
4. Results of NLP analysis (e.g., occurrence rates for pairs of ontology terms, % of issues categorized as “severe”)

Deliverable #4 is considered a stretch goal. |
| **Desired Core Competencies and Experience of Team**(Please list the desired competencies/experience/knowledge needed by the project team members that you feel will facilitate successful project execution; Examples: specific programming language experience, familiarity or expertise in certain web technologies, databases, mobile SDKs, prior classes in certain subject areas, a love of computational complexity and efficient algorithms etc.) |
| 1. Databases
2. Search engines (e.g., SOLR) and indexing
3. Web interfaces (e.g., front end for search and NLP)
4. NLP (e.g., NLTK, AlchemyAPI)
 |
| **Other Special Considerations and Project Requirements**(Please provide any special circumstances, constraints, and requirements needed by the project team members; **Examples**:* University participants must be US Citizens or Permanent Residents,
* All work is to be performed off-campus at a specific site,
* *Interdisciplinary project*: You would like to see CS students teamed with engineering students from one or more of: Computer Engineering, Electrical Engineering, Mechanical Engineering, Industrial Engineering, Civil and Environmental Engineering (please specify)
* All team members and the professor must submit to background checks,
* All team members and the professor must sign non-disclosure agreements
 |
|  |
| **Project Mentor(s), if different than who is listed above**(Please provide the contact information and title/position for the project mentor(s), who will advise the student team) |
| Professor Pawel Wocjan has offered to mentor the team on the NLP components of this project. |

*\*IMPORTANT NOTE: Proposed projects may not be chosen by student groups. In any one semester the number of potential industry-sponsored, faculty-proposed, or student-funded projects may exceed the number of student teams. If this project proposal is approved by Dr. Heinrich as a potential CS Senior Design project, you or an appropriate representative will be asked to come to class and give a 15-minute project pitch to the students. Keep in mind this is your chance to convince the students why they should pick your proposed project. Think about what is in it for them, what technologies they will get exposed to, what are the broader, enduring, and social impacts of the work, etc. If your project is chosen, you will be notified typically by the 4th week of the semester. If your project is not chosen, you will be notified in the same timeframe and if it makes sense for your timeline, we would love to offer the same project in the next semester.*