Red Team (LASI)

CS411 Spring 2013

Lab 1- Outline

**1.  Introduction**

* Introduce name (LASI) and define the terms needed to understand overview of LASI including:
  + LASI
  + Linguistic Analysis
  + Decision Support Tool
  + Importance of Themes
* Introduce Societal Problem
  + “It is difficult for people to identify a common theme over a large set of documents in a timely, consistent and objective manner.”
* Introduce Case Study elements
  + NCSOSE
  + Dr. Hester & Dr. Meyers
  + The AID process
* Describe how LASI fits into the case study and how it benefits the case study
  + Cuts down on time Dr. Hester & Dr. Meyers spend reading through documents needed for their clients individual needs
  + Allows for accurate, efficient and consistent results

**2. Product Description**

**2.1 Key Product Features and Capabilities**

* LASI Capabilities
  + Find commonalities between documents using
    - part of speech
    - statistics
* Input
  + File Types
    - txt
    - doc
    - docx
    - pdf
  + Additional documents can be added
  + User dictionary
  + Assumptions about content
* Output
  + GUI Results Display
    - Top Results
      * All documents
      * Individual documents
    - Word Relationships
      * Individual documents
    - Word Count and Weighting
      * Individual documents
  + Exportable results
    - pdf

**2.2 Hardware Platform**

* Requirements
  + CPU and Main Memory
  + Secondary Storage and Caching
* Algorithmic Document Representation
  + Word and Phrase Constructs
    - Their Instances
    - Their Roles
      * Binding
      * Referencing
    - Grammatical Abstraction Interfaces
  + Document Traversal
    - Lexical Links
    - Grammatical Links
    - Contextual Links
* Abstract View of Analysis
  + Primary
    - Lexical
      * Tokenization
      * Tag Parsing
    - Role wise Identification
      * From Tags to Runtime Constructs
        + Specialization
        + Generalization
        + Aggregation
    - Basic Metrics
  + Secondary
    - Syntactic
      * Binding
      * Inferring
    - Locally Contextual
      * Pronoun Mapping
      * Synonym Identification
    - Type Lifting
    - Intermediate Metrics
  + Tertiary
    - Semantic
      * Assumption Branching
      * Heuristic Comparison
      * Multi-pass Comparisons

**3. Identification of Case Study**

* Dr. Hester and Dr. Meyers
  + Work for NCSOSE
    - What NCSOSE does
      * Current AID Process
        + What it is
        + Currently done manually
        + Requires extensive research
        + Requires expert level knowledge
        + Bottleneck at Assessment phase
      * Reformed AID Process with LASI
        + Focus on Assessment phase
        + Assist with the analysis of documents
        + Can be used to verify Dr. Hester’s assessment
        + Can be used to defend his assertions
        + Decreases time spent analyzing documents

**4. Prototype Description**

* What it will still do
  + Accept multiple documents
  + Assign Parts of speech to words
  + Assign Parts of speech to phrases
  + Assign weights to words
  + Graphical display of results
* What it will not do
  + Be as versatile by
    - Only accepting strategic documents
      * Our case study
      * Improved accuracy
    - Only accepting doc/docx files
      * Lack of OCR tool
      * Homogeny of word recognition
    - Limiting number of documents
      * Simplified Output
      * Smaller input pool
      * Threading
      * Debugging time
    - Limiting number of pages in output
      * Simplified output
      * Smaller input pool
      * Threading
      * Debugging time
    - Giving priority to subjects, verbs, direct objects, and indirect objects
      * Fewer false positives than PoS tagging
      * Increased accuracy
      * Synonym recognition
        + Recognizable by locations in phrases

**4.1 Major Functional Components (Hardware & Software)**

**4.2. Features and Capabilities**

* Time restraints on the semester cause for the emergence of two prototypes:
  + The full prototype is our 'heads-in-the-clouds' version
    - It fully implements all of our wildest dreams
  + The real-world prototype's functionality had to be lessened
    - We limited/removed certain functions:
      * Limited types of input documents
      * Limited the length of input documents
      * Limited the number of input documents
      * Removed visual representation of logic
      * Removed synonym identification
      * Removed some user-input elements:
        + Keywords
        + Word dictionaries
        + Content assumptions
        + Specified document types
      * Reduced number of passes through document
      * Included only noun-verb binding
* [insert full prototype and real-world prototype comparison chart]