System Summary and Timing
Organization Name: Queens College, City University of New York
List of Run ID's: Pircs1,Pircs2 Pircs3,Pircs4

- Construction of Indices, Knowledge Bases, and other Data Structures
  - Methods Used to build Data Structures
    - Length (in words) of the stopword list: 630
    - Stemming Algorithm: Porters algorithm
    - Term Weighting: yes
    - Phrase Discovery? :
      - Kind of Phrase: 2-word
      - Method Used (statistical, syntactic, other): statistical
  - Statistics on Data Structures built from TREC Text
    - Special Routing Structures
      - Run ID : pircs3, pircs4
      - Type of Structure: Network of linked NODE and EDGE files capturing the query expansion terms and learnt weights
      - Automatic Process? (If not, number of manual hours): yes
      - Brief Description of Method: built from direct files of queries and documents and known relevant document information
    - Other Data Structures built from TREC text
      - Run ID : Pircs1,pircs2,pircs3,pircs4
      - Type of Structure: Compressed, truncated direct file; network of linked NODE and EDGE files built during query time
      - Total Storage (in MB): direct file - about 80MB per 500MB raw text; network - about 60MB for 10 queries per 500MB textbase
      - Total Computer Time to Build (in hours): About 5 min per 10 query
      - Automatic Process? (If not, number of manual hours): Yes
      - Brief Description of Method: built from direct files of queries and documents
  - Data Built from Sources Other than the Input Text
    - Internally-built Auxiliary File
      - Domain (independent or specific): independent
      - Type of File (thesaurus, knowledge base, lexicon, etc.): Stop words
      - Total Storage (in MB): .004
      - Number of Concepts Represented: 630
      - Type of Representation: array
      - Total Computer Time to Modify for TREC (if already built): none
    - Internally-built Auxiliary File
      - Domain (independent or specific): independent
      - Type of File (thesaurus, knowledge base, lexicon, etc.): 2-word Phrases
      - Total Storage (in MB): .005
      - Number of Concepts Represented: 528

· Type of Representation: array

- Query construction
  - Automatically Built Queries (Ad-Hoc)
    * Topic Fields Used: title, desc, narr; Boolean: title, desc
    * Average Computer Time to Build Query (in cpu seconds): 3 sec
    * Method used in Query Construction
      · Term Weighting (weights based on terms in topics)? : yes
      · Phrase Extraction from Topics? :yes, 2-word
      · Expansion of Queries using Previously-Constructed Data Structure? : yes
        I. Structure Used: Network
      · Automatic Addition of Boolean Connectors or Proximity Operators? : yes
  - Automatically Built Queries (Routing)
    * Topic Fields Used: title, desc, narr, con
    * Average Computer Time to Build Query (in cpu seconds): 18
    * Method used in Query Construction
      · Terms Selected From
        I. Topics: yes
        II. Only Documents with Relevance Judgments: yes
      · Term Weighting with Weights Based on terms in
        I. Only Documents with Relevance Judgments: yes
      · Phrase Extraction from
        I. Topics: yes
        II. Only Documents with Relevance Judgments: yes
      · Expansion of Queries using Previously-Constructed Data Structure:
        I. Structure Used: Network

- Searching
  - Search Times
    * Run ID : Pircs3,Pircs4
    * Computer Time to Search (Average per Query, in CPU seconds):about 2-3 min clock time
    * Component Times : Build network 4 min (per 10 query) Retrieval 12 min (per 10 query) Sort, merge reformat results 20 min
  - Machine Searching Methods
    * Probabilistic Model? : yes
    * Boolean Matching? : yes
    * Neural Networks? : yes
  - Factors in Ranking
    * Term Frequency? : yes
    * Other Term Weights? : yes, within-doc term frequency, inverse collection term frequency.

    * Proximity of Terms? : yes, 2 word phrases
    * Document Length? : yes
- Machine Information
    * Machine Type for TREC Experiment: Sparc 10-30
    * Was the Machine Dedicated or Shared: Dedicated
    * Amount of Hard Disk Storage (in MB): 6000
    * Amount of RAM (in MB): 128
- System Comparisons
    * Amount of "Software Engineering" which went into the Development of the System: Efficiency improvement were made cutting retrieval time by a factor of 3.
    * Given appropriate resources
        · Could your system run faster? :yes
        · By how much (estimate)? : probably half the time.
    * Features the System is Missing that would be beneficial: ability to differentiate contexts