System Summary and Timing
Organization Name: New York University
List of Run ID's: nyuir1, nyuir2

- Construction of Indices, Knowledge Bases, and other Data Structures
  - Methods Used to build Data Structures
    - Length (in words) of the stopword list: 365
    - Controlled Vocabulary? : no
    - Stemming Algorithm: yes, lexicon
      - Morphological Analysis: partial
    - Term Weighting: yes
    - Phrase Discovery? : yes
      - Kind of Phrase: syntactic
      - Method Used (statistical, syntactic, other): syntactic with statistical disambiguation
    - Syntactic Parsing? : yes
    - Proper Noun Identification Algorithm? : yes
    - Tokenizer? yes :
      - Patterns which are tokenized: names, fixed phrases
    - Other Techniques for building Data Structures: domain map
  - Statistics on Data Structures built from TREC Text
    - Inverted index
      - Run ID : nyuir
      - Total Storage (in MB): 804
      - Total Computer Time to Build (in hours): 20
      - Automatic Process? (If not, number of manual hours): yes
      - Use of Term Positions? : no
      - Only Single Terms Used? : no
    - Clusters
      - Run ID : not used in TREC-3

- Query construction
  - Automatically Built Queries (Ad-Hoc)
    - Topic Fields Used: all
    - Average Computer Time to Build Query (in cpu seconds): 2
    - Method used in Query Construction
      - Term Weighting (weights based on terms in topics)? : yes
      - Phrase Extraction from Topics? : yes
      - Syntactic Parsing of Topics? : yes
      - Word Sense Disambiguation? : no
      - Proper Noun Identification Algorithm? : yes
      - Tokenizer? : yes
        I. Patterns which are Tokenized: names, fixed phrases

- Heuristic Associations to Add Terms? : no
        - Expansion of Queries using Previously-Constructed Data Structure? : yes, but not used in TREC-3
        - Automatic Addition of Boolean Connectors or Proximity Operators? :no
  - Automatically Built Queries (Routing)
    - Topic Fields Used: desc, con
    - Average Computer Time to Build Query (in cpu seconds): 2
    - 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. Topics: no
        - II. All Training Documents: yes
      - Phrase Extraction from
        - I. Topics: yes
        - II. All Training Documents: yes
      - Syntactic Parsing
        - I. Topics: yes
        - II. All Training Documents: yes
      - Proper Noun Identification Algorithm from
        - I. Topics: yes
        - II. All Training Documents: yes

- Searching
  - Search Times
    - Run ID : nyuir
    - Computer Time to Search (Average per Query, in CPU seconds): 60
  - Machine Searching Methods
    - Vector Space Model? : yes
  - Factors in Ranking
    - Term Frequency? : yes
    - Inverse Document Frequency? : yes
    - Syntactic Clues? : phrases and names weighted differently
    - Document Length? : yes
  - Machine Information
    - Machine Type for TREC Experiment: Sun SparcStation 10
    - Was the Machine Dedicated or Shared: dedicated
    - Amount of Hard Disk Storage (in MB): 500
    - Amount of RAM (in MB): 128
  - System Comparisons
    - Amount of "Software Engineering" which went into the Development of the System: substantial, new version installed

* Given appropriate resources
  - Could your system run faster? : yes
  - By how much (estimate)? : 8-10 times
* Features the System is Missing that would be beneficial: automatic feedback