System Summary and Timing
Organization Name: U. of California, Berkeley
List of Run ID's: Brkly6, Brkly7, Brkly8

- Construction of Indices, Knowledge Bases, and other Data Structures
  - Methods Used to build Data Structures
    - Length (in words) of the stopword list: 592
    - Stemming Algorithm: SMART STEMMER
    - Term Weighting: YES
  - Statistics on Data Structures built from TREC Text
    - Inverted index
      - Run ID : Brkly6, Brkly7
      - Total Storage (in MB): 659
      - Total Computer Time to Build (in hours): APPROX. 80
      - Automatic Process? (If not, number of manual hours): YES
      - Use of Term Positions? : NO
      - Only Single Terms Used? : YES
    - Inverted index
      - Run ID : Brkly8
      - Total Storage (in MB): 364
      - Total Computer Time to Build (in hours): APPROX. 40
      - Automatic Process? (If not, number of manual hours): YES
      - Use of Term Positions? : NO
      - Only Single Terms Used? : YES
- Query construction
  - Automatically Built Queries (Ad-Hoc)
    - Topic Fields Used: TITLE, DESC, NARR
    - Average Computer Time to Build Query (in cpu seconds): APPROX. 3 / QUERY
    - Method used in Query Construction
      - Term Weighting (weights based on terms in topics)? : YES
      - Other: ABSOLUTE QUERY STEM FREQUENCIES, QUERY LENGTHS, AND IDF'S WERE USED IN WEIGHTING THE QUERY STEMS
  - Automatically Built Queries (Routing)
    - Topic Fields Used: DOM, TITLE, DESC, NARR, CON, DEF, NAT TIME
    - Average Computer Time to Build Query (in cpu seconds): APPROX. 100 / QUERY
    - 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: YES

II. Only documents with Relevance Judgments: YES

- Manually Constructed Queries (Ad-Hoc)
  - Topic Fields Used: TITLE, DESC, NARR
  - Average Time to Build Query (in Minutes): 20-30
  - Tools used to Build Query
    - Other Lexical Tools? : Thesaurus allowed but not used much.
  - Method used in Query Construction
    - Addition of Terms not Included in Topic? : YES
      I. Source of Terms: Boolean lookups in parallel collections

- Searching
  - Search Times
    - Run ID : Brkly6, Brkly7
    - Computer Time to Search (Average per Query, in CPU seconds): APPROX 16 SECONDS PER QUERY TO SEARCH ANY ONE OF THE FIVE COLLECTIONS SEPARATELY.
    - Run ID : Brkly8
    - Computer Time to Search (Average per Query, in CPU seconds): APPROX. 300
  - Machine Searching Methods
    - Machine Searching Methods
      - Probabilistic Model? : YES
  - Factors in Ranking
    - Factors in Ranking
      - Term Frequency? : YES
      - Inverse Document Frequency? : YES
      - Document Length? : YES
  - Machine Information
    - Machine Type for TREC Experiment: DEC STATION 5000/125
    - Was the Machine Dedicated or Shared: Dedicated
    - Amount of Hard Disk Storage (in MB): 4 GB
    - Amount of RAM (in MB): 48
    - Clock Rate of CPU (in MHz): 15 MHZ
  - System Comparisons
    - Amount of "Software Engineering" which went into the
      - Development of the System:
        I. NONE EXCEPT FOR THE PROBABILISTIC LOGIC. THE BERKELEY SYSTEM IS AN EXPERIMENTAL PROTOTYPE ONLY, PROGRAMMED AS A MINIMAL MODIFICATION OF THE SMART SYSTEM (VERSION 10).
    - Given appropriate resources
      - Could your system run faster? : NATURALLY