System Summary and Timing
Organization Name: Verity Inc.
List of Run ID's: TOPIC1, TOPIC2, TOPIC3, TOPIC4

- Construction of Indices, Knowledge Bases, and other Data Structures
  - Methods Used to build Data Structures
    - Length (in words) of the stopword list: No content stop words used in building the collections; some reserved words (i.e., SGML tags)removed
    - Stemming Algorithm: Verity proprietary, but similar in spirit to Porter
      - Morphological Analysis: Limited to that provided by the stemmer
    - Tokenizer? : No special tokens recognized for TREC
  - Statistics on Data Structures built from TREC Text
    - Inverted index
      - Run ID : all runs
      - Total Storage (in MB): 1,504
      - Total Computer Time to Build (in hours): 112 hrs(elapsed time)
      - Automatic Process? (If not, number of manual hours): Yes
      - Use of Term Positions? : Yes
      - Only Single Terms Used? : Yes
- Query construction
  - Interactive Queries
    - Initial Query Built Automatically or Manually: Manually for routing test; automatically for ad hoc test
    - Type of Person doing Interaction
      - System Expert: Some; we used a team of query builders, some of whom were proficient in the use of TOPIC
    - Average Time to do Complete Interaction
      - CPU Time (Total CPU Seconds for all Iterations): No data
      - Clock Time from Initial Construction of Query to Completion of Final Query (in minutes): 116 mins (estimate) averaged across routing and adhoc queries; routing queries on average somewhat lower, adhoc queries on average somewhat higher
    - Average Number of Iterations: No data
    - Average Number of Documents Examined per Iteration: No data
    - Minimum Number of Iterations: No data
    - Maximum Number of Iterations: No data
    - What Determines the End of an Iteration: the point at which the query is rerun and the results rescored
    - Methods used in Interaction
      - Manual Methods
        I. Using Individual Judgment (No Set Algorithm)? : Yes
- Searching
  - Search Times

    * Run ID : averaged over all runs
    * Computer Time to Search (Average per Query, in CPU seconds): no data; average elapsed time is 3 mins (est.) but this varies widely with factors such as query complexity, field search constraints, data subsets being used, network load, etc.

    * Component Times : no separate data available, but times are dominated by query matching and scoring
- Machine Searching Methods
    * Other: Verity's concept-based retrieval technology which uses a mixture of knowledge-based and probabilistic techniques
- Factors in Ranking
    * Term Frequency? : yes for some operators
    * Semantic Closeness? : yes
    * Proximity of Terms? : yes for some operators
    * Percentage of Query Terms which match? : yes for some operators
- Machine Information
    * Machine Type for TREC Experiment: SPARC 2 used as file server
    * Was the Machine Dedicated or Shared: Dedicated for TREC, but shared by all people doing the query building and experiments
    * Amount of Hard Disk Storage (in MB): 4600
    * Amount of RAM (in MB): 32MB
    * Clock Rate of CPU (in MHz): Unknown
- System Comparisons
    * Amount of "Software Engineering" which went into the Development of the System: TOPIC is COTS
    * Given appropriate resources
        · Could your system run faster? : of course
        · By how much (estimate)? : hard to tell
    * Features the System is Missing that would be beneficial: a capability that allowed us to focus on specific "zones" of documents would probably reduce the over generation problem we have