System Summary and Timing
Organization Name: West Publishing Co.
List of Run ID's: westp1 (Ad Hoc) westp2 (Routing)

- Construction of Indices, Knowledge Bases, and other Data Structures
  - Methods Used to build Data Structures
    - Length (in words) of the stopword list: 290
    - Stemming Algorithm: Modified Porter Stemmer
      - Morphological Analysis: Yes. There are 2789 words in our exception list. The remaining words are stemmed using the Porter Algorithm.
    - Term Weighting: Yes.
    - Phrase Discovery? : Yes.
      - Kind of Phrase: File of common phrases. (approx. 32000).
    - Tokenizer? : A small collection of synonyms for company names, countries, states, and months is used.
  - Statistics on Data Structures built from TREC Text
    - Inverted index
      - Total Storage (in MB): westp1 (adhoc) 523
        westp2 (routing) 380

      - Total Computer Time to Build (in hours):
        westp1 (adhoc) 42 hrs.
        westp2 (routing) 16 hrs.

      - Automatic Process? (If not, number of manual hours):
        westp1 (adhoc) yes
        westp2 (routing) yes

      - Use of Term Positions? : westp1 (adhoc) yes
        westp2 (routing) yes

      - Only Single Terms Used? : westp1 (adhoc) yes
        westp2 (routing) yes

- Query construction
  - Automatically Built Queries (Ad-Hoc)
    - Topic Fields Used: title, desc, narr
    - Average Computer Time to Build Query (in cpu seconds): .5 secs per query. (25 seconds for 50 queries.)
    - Method used in Query Construction
      - Term Weighting (weights based on terms in topics)? : yes.
      - Phrase Extraction from Topics? : yes.
      - Tokenizer? :
        I. Patterns which are Tokenized: some company and country names.

- Automatically Built Queries (Routing)
  - Topic Fields Used: title, con
  - Method used in Query Construction
    - Terms Selected From
      I. Topics: yes
    - Term Weighting with Weights Based on terms in
      I. Topics: yes. weight of 2 given to terms in title
    - Tokenizer: yes. some acronyms such as "ASK" and "RISC" were recognized.

- Searching
  - Search Times
    - Run ID: westp1
    - Computer Time to Search (Average per Query, in CPU Seconds): Averaged 349 secs/query (clock time).
  - Machine Searching Methods
    - Probabilistic Model? : Yes.
  - Factors in Ranking
    - Term Frequency? : Yes.
    - Inverse Document Frequency? : Yes.
    - Other Term Weights? : Yes. Differential weights were assigned.
  - Machine Information
    - Machine Type for TREC Experiment: SUN Sparc 5
    - Was the Machine Dedicated or Shared: Shared.
    - Amount of RAM (in MB): 64
    - Clock Rate of CPU (in MHz): 70
  - System Comparisons
    - Amount of "Software Engineering" which went into the Development of the System: A couple of years.
    - Given appropriate resources
      - Could your system run faster? : Yes.