

clude collections of documents and collections of search requests, and the relevant documents available in the collection are known for each of these search requests. Hence, all systems participating in the experiment should use both documents and search requests from the benchmark data.

The collection of documents used in these experiments consists of approximately 1 million documents. The selection of such a large set is justified as follows: "Evaluation using the small collections currently available may not reflect performance of systems in large full-text searching" (Harman, 1995). In other words, it is assumed that in a smaller collection (say, of 100,000 documents), experiment results would be distorted. It should be noted that in the famous Cranfield experiments, the collection consisted of 1400 documents. Be that as it may, to participate in the experiment it is necessary first to index approximately 1 million documents and then to perform the search.

Another part of the benchmark was described as follows:

retrieving lists of documents that could be considered relevant to each of the 50 topics in what was called "ad hoc" querying. A second information retrieval paradigm used was where 50 retrieval topics were known in advance and new documents were to be matched against the 50 standard queries simulating a "routing" operation. In both cases the queries were not really queries at all but carefully honed user need statements and were thus extensive descriptions of the topic of interest. (Report on TREC-2, 1993)

In other words the experiments are conducted not with real search requests and real users (output is evaluated by nonusers) but with specially matched topics and evaluations of the relevance of documents performed by experts.

Organizers of the experiment cite the following example of a topic and a document:

```
<top>
Attachment 1—Sample topic
<head> Tipster Topic Description
<num> Number: 028
<dom> Domain: Science and Technology
<title> Topic: AT&T Technical Efforts
<desc> Description: Document must describe AT&T's technical efforts in computers and communications.
```

```
<nar> Narrative: To be relevant, a document must contain information on the efforts of American Telephone and Telegraph (AT&T) in computers and communications. Examples of relevant subject matter would include product announcements, releases or cancellations, and a discussion of AT&T Bell Labs research. Documents focusing on either AT&T's efforts
```

to buy other computer companies, AT&T's legal battles with other organizations, or AT&T's Unix operating system are not relevant. For the purposes of this topic, the Regional Bell Operating Companies (RBOCs) or "Baby Bells" are not considered AT&T.

```
<con> Concept(s):
```

1. AT&T, American Telephone and Telegraph
2. 3B-2 minicomputer, AT&T 386 PC
3. AT&T Starlan
4. PBX
5. Product announcements, product releases

```
</top>
```

```
Attachment 2—Sample document (abridged)
```

```
<DOC>
```

```
<DOCNO> WSJ880460-0090 </DOCNO>
```

```
<HL> AT&T Unveils Services to Upgrade Phone Networks under Global Plan
```

```
</HL>
```

```
<AUTHOR> Janet Guyon (WSJ Staff) </AUTHOR>
```

```
<SO> </SO>
```

```
<CO> T </CO>
```

```
<IN> TEL </IN>
```

```
<DATELINE> NEW YORK </DATELINE>
```

```
<TEXT> American Telephone and Telephone Company introduced the first of a new generation of phone services with broad applications for computer and communications equipment makers.
```

AT&T said it is the first national long-distance carrier to announce prices for specific services under a worldwide plan to upgrade phone networks. By announcing commercial services under the plan, which the industry calls the Integrated Services Digital Network, AT&T will influence evolving communications standards to its advantage, consultants said, just as International Business Machines Corp. has created de facto computer standards favoring its products.

```
</TEXT>
```

```
</DOC>
```

It should be noted that the Cranfield experiment involved 225 real search requests (not topics) and a relevancy evaluation was performed by users. Con-