ACSys/CSIRO TREC-9 Experiments

David Hawking*
CSIRO Mathematics and Information Sciences,
Canberra, Australia
David.Hawking@cmis.csiro.au

February 7, 2001

Unfortunately, ACSys was able to commit very few resources to TREC experiments this year and participated only in the Web track. I took the retrieval component (PADRE99) of our standard metadata/content Intranet search engine, modified it to handle TREC-formatted web pages and ran the unexpanded topic titles as queries. I also generated a set of manual (non-interactive) queries from the topic statements.

It transpired that the query "angioplast7" generated by the automatic title-only process retrieved 0 documents and these runs were rejected by the NIST submission process. Therefore, only the manual run (`acsys9mw0`) was officially assessed. It achieved an average precision of 0.2486.

As in TRECs 6, 7 and 8 [Hawking et al. 1997], the basic relevance scoring method used in official ACSys adhoc runs the Okapi BM25 weighting function [Robertson et al. 1994]

$$w_t = q_t \times tf_d \times \frac{\log(\frac{N-n+0.5}{n+0.5})}{2 \times (0.25 + 0.75 \times \frac{dl}{avdl}) + tf_d} \tag{1}$$

where w_t is the relevance weight assigned to a document due to query term t, q_t is the weight attached to the term by the query, tf_d is the number of times t occurs in the document, N is the total number of documents, n is the number of documents containing at least one occurrence of t, dl is the length of the document and $avdl$ is the average document length (measured either in bytes or in indexable words). A slight modification has been applied to ensure that weights are never negative.

As in TREC-8 runs using PADRE99, the index included stopwords, numbers and strings comprised of letters and digits. Words in the index were not stemmed but query-time stemming and truncation operators were available. In the official (manual) run, no automatic expansion or relevance feedback was employed.

The queries listed on the next page were processed as follows. In general, each query element (word, stem, phrase or dysjunctive group) was treated as a constraint. Regardless of BM25 score, documents matching more constraints are presented ahead of documenents matching fewer constraints. This results in a tiered ranking. Within each tier, documents are ranked by BM25 score.

*The author wishes to acknowledge that this work was carried out within the Cooperative Research Centre for Advanced Computational Systems established under the Australian Government's Cooperative Research Centres Program. Despite this, the work was actually carried out in Greece, while attending a Summer School and a Conference.

Example	Feature	Notes
"babe ruth"	Phrase	
theater#"	Stem operator	
[pine pinus]	Dysjunctive group	Constraint satisfied by either word.
~grow	Non-constraint operator	Word contributes to score but is not a constraint.
recurr*	Truncation operator	Matches any word starting with recurr

```
451 Bengal [cat cats]
452 [beaver beavers] [salt saline fresh saltwater freshwater sea seawater estuary]
453 [hunger famine malnutrition starvation starving] [relief aid feed*]
454 ["parkinson's disease" parkinsonism]
455 "jackie robinson" ["first game" "first appearance" "debut"] [year season]
456 ["world end" "end of the world" apocalypse] 2000 [group sect cult religion followers disciples]
457 chevrolet [truck trucks]
458 [fast fasting] [religion islam muslim jewish judaism catholic*]
459 [foreclose foreclosed foreclosure]
460 Moses ~[israel* prophet "red sea" "burning bush" commandments egypt pharoah]
461 "lava lamp#"
462 ["real estate" realty realtor "estate agent#"]["new jersey" NJ] ~[house houses appartment# residen*]
463 [tartan tartans plaid plaids] ~[scottish clan clans]
464 "nativity scene*" [ban bans banning banned]
465 [disease diseases syndrome] [deer venison]  ~[lyme tick ticks]
466 "peer gynt" ~[compose grieg composer suite folk folklore]
467 [dachshund dachshunds "wiener dog"] ~[breed breeder pedigree]
468 incandescent [light bulb# globe# filament tungsten] [carbon edison swann invent*]
469 steinbach nutcracker*
470 mistletoe [beneficial benefit health medical medicine drug nutrition feed food sustain cure]
471 "mexican food" [australia* europe* asia asian afric* spain philipines]
472 antique appliance restor* ~[dealer* collect* museum*]
473 "toronto film" [award won winner winning prize]
474 ["e mail" email e-commerce] [profit# turnover sales money] ~[internet online dotcom]
475 zirconium [melting boiling hardness valency *valent react reaction ductil* refractory malleable "atomic number"]
476 "jennifer aniston" [movie movies cinema tv television]
477 [rccl "royal carribean"] [ship ships liner liners vessel vessels]
478 baltimore mayor*
479 "kappa alpha psi" [fraternity sorority college# university#]
480 [traffic roads cars interstate vehicles] [washington DC maryland virginia arlington bethesda beltway]
481 "babe ruth" 1920
482 "growth rate" [pine pinus pinetree pines] ~[grow growing]
483 [rosebowl "rose bowl"] parade
484 [auto car cars vehicle#] skoda
485 [gps "global positioning"] clock accuracy
486 [eldorado "el dorado"] reno ~[nevada casino street road]
487 angioplasty [repeat follow-up recurr*]
488 "newport beach" [california ca] [entertain* movies arcade* sport# game games theater#]
489 calcium [diet dietary supplement# health osteoporosis bone]
490 [motorcycle motorbike] helmet# [law legal violation compulsory]
491 ["japanese wave" tsunami] ~[damage killed died property destroyed "swept away" relief aid]
492 "savings bond#" ~[us federal interest maturity rate* denominat*]
493 [retirement communit*]
494 nirvana ~[guitar* drum* lyrics penned songs "rock group" "rock band"]
495 twenties
496 [tmj "temporal mandible"] [syndrome condition suffer*] ~[treatment therapy symptom* cause* "due to"]
497 [orchids "orchid growing"]
498 "hair transplant#" ~[follicle* bald* scalp]
499 "pool cue"
500 ["dna test*" "dna analysis"] ~[forensic police fbi investigat* dna]
```

Bibliography

Hawking, D., Thistlewaite, P., and Craswell, N. 1997. ANU/ACSys TREC-6 experiments. In E. M. Voorhees and D. K. Harman Eds., *Proceedings of TREC-6* (Gaithersburg MD, November 1997), pp. 275–290. NIST special publication 500-240, `http://trec.nist.gov`.

Robertson, S. E., Walker, S., Hancock-Beaulieu, M., and Gatford, M. 1994. Okapi at TREC-3. In D. K. Harman Ed., *Proceedings of TREC-3* (Gaithersburg MD, November 1994). NIST special publication 500-225.