

An Information Retrieval Application for Simulated Annealing

Major B. J. Jansen

Dept. of Electrical Engineering and Computer Science
United States Military Academy
West Point, New York 10996
Telephone: (914) 938-3233
FAX: (914) 938-5956
jjansen@acm.org

Dr. Udo Pooch

Computer Science Department
H. R. Bright Building
Texas A&M University
College Station, Texas 77843-3112
Telephone: (409)845-5498
FAX: (409) 847-8578

Please Cite: Jansen, B. J. 1997. An information retrieval application for simulated annealing. The 2nd ACM Conference on Digital Libraries. Philadelphia, PA., 259-260.

[See Other Publications](#)

With information retrieval, there is little time and limited information available to build an extensive profile of user desires. To improve information retrieval performance, one must rapidly model user desires using the best information available. One must also change the model if better information becomes available. With a search engine, this means that one must rank and quickly change the ranking of query results to meet changes in user desires. While relevance feedback algorithms are effective in building a user model, techniques to rapidly change the model are less successful. This research combines a modified relevance feedback with a simulated annealing algorithm to achieve this rapid remodeling.

Unlike relevance feedback, simulated annealing is not widely used in information retrieval. A simulated annealing algorithm has both greedy (deterministic) and random (stochastic) characteristics. The deterministic aspect attempts to improve upon the current state using a predefined cost function. However, the stochastic aspect occasionally accepts a state that is NOT an improvement. Simulate annealing has proven an effective technique for problems where there are many solutions and currently no way to derive which is the best one.

With these types of problems, one must test and compare possible solutions to find a near-optimal one. A deterministic algorithm always attempts to improve on the current state and may become "stuck" in a local solution. A local solution is a state that is not the best solution but where any change from the current state increases the cost. A stochastic

algorithm takes too long to arrive at the optimal solution. Simulated annealing overcomes both the local solution and the time problem. This situation of many solutions and no way to calculate the best one is analogous to finding the optimal ranking of query results.

With a search engine, the relevance feedback algorithm attempts to reformulate the query based on characteristics of relevant documents. Most search engines then generate a new query result list. Research has shown that relevance feedback improves recall and precision for the reformulated query. Unfortunately, the reformulated query may no longer represent the information desired. For example, the user may have gotten sidetracked (i.e., searching on computers but browsing the Bill Gates homepage) or the user may desire a broad search (i.e., scholarly journals exclusively on computer architecture, scholarly journals related to computer architecture, and popular press journals on computer architecture). Many relevance feedback algorithms, such as Excite's *More_Like_This*, cannot handle broad searches.

This research is exploring a relevance feedback algorithm to rerank a query result list rather than generating a new result list. Naturally, one must remove documents a user viewed from the next ranking. The original query results list always remains accessible to the user. The user can then explore a broad subject without restarting the search process. A problem has been switching between subtopics. It is difficult to overcome the effect of existing information. Simulated annealing's stochastic characteristic may rapidly overcome previous information concerning document relevance. With this ability, simulated annealing may make search engines better information retrieval assistants.