Search Engine: Iteration 2

Ranking

> All search results are equal, but some search results are more equal than others.
>
> *George Orwell, Animal Farm*

Having successfully laid the groundwork for our local search engine with robust indexing in the first iteration, we now turn our attention to ranking. The most basic signal that information is relevant is when content contains the same keywords as your search query. More sophisticated ranking techniques take into account the length of the documents and assign a numeric *weight* to each term.

Although your DBMS handles a lot of content-based ranking, you have enough metadata to make the search results even more relevant. Some factors that can boost ranking include:

- Path length
- Keyword presence in path
- Directory importance
- File extension prioritization
- Recent file access
- File size

Search Engine: Iteration 2

Software Design 2025
March 31, 2025

Score	Criteria
5/10	**(Passing - Basic Functionality):** Implement a query parser that can process a query like `path:A/B content:C`. Your basic language should support any permutation of these qualifiers. Duplicate qualifiers should be combined using the `AND` operator. As a precondition, you MUST index your real file system content. You will not get a passing grade if: • You demo your application on a limited test set of 10 files. • I manage to search and understand your codebase faster than you do.
6/10	**(Below Average):** Implement a function that scores file paths at index time. These weights will determine the order in which your search results are displayed at query time.
7/10	**(Average):** Add a new index report format that can be configured at start-up time.
8/10	**(Good):** Track your search activity. Consider using the Observer pattern for this. Use the search history to: • Suggest queries. • Rank results.
9/10	**(Excellent):** Your own sufficiently-complex feature (approved by your TA).
10/10	**(Outstanding):** Adherence to good software engineering practices (aforementioned principles from iteration 1, parser/ranking tests, etc.).

Important: The deadline for this iteration is Week 8.