

Assignment 2: Information Indexing and Retrieval Procedures

Published

Assign To

Edit

:

The purpose of this assignment is to explore the principles of designing information indexing and retrieval procedures. You will analyze a web-based information retrieval system and evaluate indexing and retrieval procedures based on a set of standards.

Below are two web-based information services. Please explore and evaluate the two search engines by applying the below principles.

1. **DTSearch Engine** [\(Textual Information\)](https://search.dtsearch.com/forms/faq/search.html)
2. **Wayfair** [\(Picture Information\)](https://wayfair.com/)

Apply the principles outlined below to analyze the following dimensions of the search engine. Your submission should include findings from all five perspectives, supported by evidence (images can be included if necessary). Each perspective is worth 4 points, with two points assigned to each of two search engines. 5 points are for overall completeness and quality of the submission.

Guidelines for Submission:

- Clearly describe your methods and results.
- Even if content overlaps between perspectives, provide a separate evaluation for each perspective to facilitate grading. Points will be deducted if evaluations are not properly organized under each perspective.

Evaluation Perspectives:

1. Indexing Types and Methods

Investigate the types of indexing used by the search engine, such as keywords, phrases, or subject headings. Additionally, identify the indexing methods employed:

- **Automatic Indexing:** May rely on keyword-based searching, recognizable through operators like AND, OR, NOT, and by its focus on keyword relevance.
- **Manual Indexing:** Could support subject headings or categorization, producing search results that include subject or category labels.
- **Metadata-Generated Indexing:** Often uses metadata fields like "author:" or "date:" and may provide facets for filtering search results.

Note: When it's challenging to identify the method directly, make educated guesses by combining information from search engine documentation or inferred from search results.

2. Search Result Relevance and Indexing Precision/Recall

Assess the relevance of search results for different query types (keywords, phrases, and subject headings). Calculate and compare:

- **Precision:** The proportion of relevant documents among the retrieved ones.
- **Recall:** The proportion of relevant documents retrieved out of all available relevant documents.

Compare the performance of various indexing types in terms of relevance and accuracy.

3. Advanced Search Features

Test whether the search engine supports advanced search functionalities such as:

- **Boolean Operators** (AND, OR, NOT)
- **Wildcard Search** (e.g., * or ?)
- **Proximity Search** (searching for terms within a specific distance of each other)

Report how effectively these features enhance query precision and recall.

4. Information Retrieval Methods

Identify the information retrieval methods employed by the two search engines. Common methods include:

- **Free Text Retrieval:** Allows natural language queries without strict syntax.
- **Boolean Logic:** Uses logical operators for precise query construction.
- **Fuzzy Logic:** Supports partial matching and measures similarity rather than strict accuracy.
- **Natural Language Processing (NLP):** Understands user intent and semantic relationships, improving result relevance.

5. Advanced Techniques

Determine if the search engines implement advanced techniques such as:

- **Stemming:** Reduces words to their root form to improve recall.
- **Phonetic Searching:** Finds words that sound similar to the query term.
- **Fuzzy Search:** Allows approximate matching, accommodating variations like typos or alternative spellings.

Additionally, evaluate features like:

- **Search Result Ratings:** Rate documents based on relevance (e.g., 1–5 scale).
- **Relevance Feedback:** Examine how user interactions (e.g., clicks, views) influence subsequent search results to improve relevance.

Deliverables:

Provide a detailed report of your findings for each perspective. Include the following:

- **Procedures:** Step-by-step methods used for your evaluation.
- **Results:** Key findings for each perspective, supported by data and screenshots if applicable.

Some Guidance for the Assignment

The most important thing to keep in mind is that I know this is new content and a bit complex, to many of you these are entirely new concepts. I am not looking to be overly punitive with my grading. If you make an attempt to address the questions, you will receive the full points. Just do the best that you can with the content and show effort and you will earn full points. So, please do not sweat being perfect in your results. In many cases, there is not a "right" answer, because the results you receive from a search can vary based on your query. That said, here is some more "plain-language" suggestions for each of the five perspectives in order to get you through the assignment:

1. First, enter a general search term in both databases and compare the structure of the results. Do you notice major differences in what information is presented to you (e.g., features like dates, hits, description). What could this tell you about the differences between the two systems?
2. Enter search terms and then calculate from the top "x" number of results how many are relevant to what you were searching and how many are not. This is your "precision" score. Recall can be a bit more tricky, but think about your own knowledge of what resources are out there on a topic compared to what you are seeing from the results of this search. Do you know of certain resources that are missing from this database's results? Then the recall is probably pretty low.
- 3-5. Test the suggested operators in your databases and see how it changes the results (if at all). Not all methods suggested in steps 3-5 will be available for both of the databases - the databases are distinct, with dtSearch providing more detailed querying features and Wayfair providing a better user interface. Simply, try the searches based on the suggested approaches and see if they generate different/improved results.

If you encounter any questions or challenges, please do not hesitate to reach out to me!

Points 25

Submitting a file upload

File Types doc, docx, pdf, and txt

Due	For	Available from	Until
Mar 1, 2026	Everyone	Feb 2, 2026 at 12am	-

[+ Create Rubric](#)[🔍 Find Rubric](#)