

Google's Query Fan-Out Patent: Thematic Search via @sejournal, @martinibuster

A patent that Google filed in December 2024 presents a close match to the Query Fan-Out technique that Google's AI Mode uses. The patent, called Thematic Search, offers an idea of how AI Mode answers are generated and suggests new ways to think about content strategy.

Disclaimer: Virtually all of the patents that Google files list multiple ways to implement them as part of an algorithm, which is also the case with this patent. Google generally does not confirm that the inventions in a patent or research paper are in use, which is the case with this patent.

What The Patent Describes

The patent describes a thematic search system that organizes related search results to a search query into categories, what it calls themes, and provides a short summary for each theme so that users can understand the answers to their questions without having to click a link to all of the different sites.

The patent describes a system for deep research, for questions that are broad or complex. What's new about the invention is how it automatically identifies themes from the traditional search results and uses an AI to generate an informative summary for each one using both the content and context from within those results.

Thematic Search Engine

Themes is a concept that goes back to the early days of search engines, which is why this patent caught my eye a few months ago and caused me to bookmark it.

Here's the TL/DR of what it does:

- The patent references its use within the context of a large language model and a summary generator.
- It also references a thematic search engine that receives a search query and then passes that along to a search engine.
- The thematic search engine takes the search engine results and organizes them into themes.
- The patent describes a system that interfaces with a traditional search engine and uses a large language model for generating summaries of thematically grouped search results.
- The patent describes that a single query can result in multiple queries that are based on "sub-themes"

Comparison Of Query Fan-Out And Thematic Search

The system described in the patent mirrors what Google's documentation says about the Query Fan-Out technique.

Here's what the patent says about generating additional queries based on sub-themes:

"In some examples, in response to the search query 142-2 being generated, the thematic search engine 120 may generate thematic data 138-2 from at least a portion of the search results 118-2. For example, the thematic search engine 120 may obtain the search results 118-2 and may generate narrower themes 130 (e.g., sub-themes) (e.g., "neighborhood A",

“neighborhood B”, “neighborhood C”) from the responsive documents 126 of the search results 118-2. The search results page 160 may display the sub-themes of theme 130a and/or the thematic search results 119 for the search query 142-2. The process may continue, where selection of a sub-theme of theme 130a may cause the thematic search engine 120 to obtain another set of search results 118 from the search engine 104 and may generate narrower themes 130 (e.g., sub-sub-themes of theme 130a) from the search results 118 and so forth.”

Here’s what Google’s documentation says about the Query Fan-Out Technique:

“It uses a “query fan-out” technique, issuing multiple related searches concurrently across subtopics and multiple data sources and then brings those results together to provide an easy-to-understand response. This approach helps you access more breadth and depth of information than a traditional search on Google.”

The system described in the patent resembles what Google’s documentation says about the Query Fan-Out technique, particularly in how it explores subtopics by generating new queries based on themes.

Summary Generator

The summary generator is a component of the thematic search system. It’s designed to generate textual summaries for each theme generated from search results.

This is how it works:

- The summary generator is sometimes implemented as a large language model trained to create original text.
- The summary generator uses one or more passages from search results grouped under a particular theme.

- It may also use contextual information from titles, metadata, surrounding related passages to improve summary quality.
- The summary generator can be triggered when a user submits a search query or when the thematic search engine is initialized.

The patent doesn't define what 'initialization' of the thematic search engine means, maybe because it's taken for granted that it means the thematic search engine starts up in anticipation of handling a query.

Query Results Are Clustered By Theme Instead Of Traditional Ranking

The traditional search results, in some examples shared in the patent, are replaced by grouped themes and generated summaries. Thematic search changes what content is shown and linked to users. For example, a typical query that a publisher or SEO is optimizing for may now be the starting point for a user's information journey. The thematic search results leads a user down a path of discovering sub-themes of the original query and the site that ultimately wins the click might not be the one that ranks number one for the initial search query but rather it may be another web page that is relevant for an adjacent query.

The patent describes multiple ways that the thematic search engine can work (I added bullet points to make it easier to understand):

- "The themes are displayed on a search results page, and, in some examples, the search results (or a portion thereof) are arranged (e.g., organized, sorted) according to the plurality of themes. Displaying a theme may include displaying the phrase of the theme.

- In some examples, the thematic search engine may rank the themes based on prominence and/or relevance to the search query.
- The search results page may organize the search results (or a portion thereof) according to the themes (e.g., under the theme of ‘cost of living”, identifying those search results that relate to the theme of ‘cost of living”).
- The themes and/or search results organized by theme by the thematic search engine may be rendered in the search results page according to a variety of different ways, e.g., lists, user interface (UI) cards or objects, horizontal carousel, vertical carousel, etc.
- The search results organized by theme may be referred to as thematic search results. In some examples, the themes and/or search results organized by theme are displayed in the search results page along with the search results (e.g., normal search results) from the search engine.
- In some examples, the themes and/or theme-organized search results are displayed in a portion of the search results page that is separate from the search results obtained by the search engine.”

Content From Multiple Sources Are Combined

The AI-generated summaries are created from multiple websites and grouped under a theme. This makes link attribution, visibility, and traffic difficult to predict.

In the following citation from the patent, the reference to “unstructured data” means content that’s on a web page.

According to the patent:

“For example, the thematic search engine may generate themes from unstructured data by analyzing the content of the

responsive documents themselves and may thematically organize the search results according to the themes.

...In response to a search query (“moving to Denver”), a search engine may obtain search results (e.g., responsive documents) responsive to that search query.

The thematic search engine may select a set of responsive documents (e.g., top X number of search results) from the search results obtained by the search engine, and generate a plurality of themes (e.g., “neighborhoods”, “cost of living”, “things to do”, “pros and cons”, etc.) from the content of the responsive documents.

A theme may include a phrase, generated by a language model, that describes a theme included in the responsive documents. In some examples, the thematic search engine may map semantic keywords from each responsive document (e.g., from the search results) and connect the semantic keywords to similar semantic keywords from other responsive documents to generate themes.”

Content From Source Pages Are Linked

The documentation states that the thematic search engine links to the URLs of the source pages. It also states that the thematic search result could include the web page’s title or other metadata. But the part that’s important for SEOs and publishers is the part about attribution, links.

“...a thematic search result 119 may include a title 146 of the responsive document 126, a passage 145 from the responsive document 126, and a source 144 of the responsive document. The source 144 may be a resource locator (e.g., uniform resource location (URL)) of the responsive document 126.

The passage 145 may be a description (e.g., a snippet obtained from the metadata or content of the responsive document 126). In some examples, the passage 145 includes a portion of the responsive document 126 that mentions the respective theme 130. In some examples, the passage 145 included in the thematic search result 119 is associated with a summary description 166 generated by the language model 128 and included in a cluster group 172.”

User Interaction Influences Presentation

As previously mentioned, the thematic search engine is not a ranked list of documents for a search query. It’s a collection of information across themes that are related to the initial search query. User interaction with those AI generated summaries influences which sites are going to receive traffic.

Automatically generated sub-themes can present alternative paths on the user’s information journey that begins with the initial search query.

Summarization Uses Publisher Metadata

The summary generator uses document titles, metadata, and surrounding textual content. That may mean that well-structured content may influence how summaries are constructed.

The following is what the patent says, I added bullet points to make it easier to understand:

- “The summary generator 164 may receive a passage 145 as an input and outputs a summary description 166 for the inputted passage 145.

- In some examples, the summary generator 164 receives a passage 145 and contextual information as inputs and outputs a summary description 166 for the passage 145.
- In some examples, the contextual information may include the title of the responsive document 126 and/or metadata associated with the responsive document 126.
- In some examples, the contextual information may include one or more neighboring passages 145 (e.g., adjacent passages).
- In some examples, the contextual information may include a summary description 166 for one or more neighboring passages 145 (e.g., adjacent passages).
- In some examples, the contextual information may include all the other passages 145 on the same responsive document 126. For example, the summary generator may receive a passage 145 and the other passages 145 (e.g., all other passages 145) on the same responsive document 126 (and, in some examples, other contextual information) as inputs and may output a summary description 166 for the passage 145."

Thematic Search: Implications For Content & SEO

There are two way that AI Mode ends for a publisher:

1. Since users may get their answers from theme summaries or dropdowns, zero-click behavior is likely to increase, reducing traffic from traditional links.
2. Or, it could be that the web page that provides the end of the user's information journey for a given query is the one that receives the click.

I think this means that we really need to re-think the paradigm of ranking for keywords and maybe consider what the question is that's being answered by a web page, and then identify follow-up questions that may be related to that

initial query and either include that in the web page or create another web page that answers what may be the end of the information journey for a given search query.

You can read the patent here:

Thematic Search (PDF)

Read Google's Documentation Of AI Mode (PDF)

<https://www.searchenginejournal.com/google-query-fan-out-patent/547983/>