An architecture for non-Linear discovery of aggregated multimedia document web search results (#54923)

First submission

Guidance from your Editor

Please submit by **30 Dec 2020** for the benefit of the authors (and your $200 publishing discount) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

4 Figure file(s)
2 Latex file(s)
1 Raw data file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	***Example***
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

An architecture for non-Linear discovery of aggregated multimedia document web search results

Umer Rashid Corresp., 1 , **Abdur Khan** Corresp., 1 , **Khalid Saleem** [1] , **Adeel Ahmed** [1]

[1] Department of Computer Science, Quaid-i-Azam University, Islamabad, Islamabad, Pakistan

Corresponding Authors: Umer Rashid, Abdur Khan
Email address: umerrashid@qau.edu.pk, arkhan@cs.qau.edu.pk

The recent proliferation of multimedia information on the web enhances user information need from simple textual lookup to multi-modal exploration activities. The current search engines act as major gateways to access the immense amount of multimedia data. However, access to the multimedia content is provided by aggregating disjoint multimedia search verticals. The aggregation of the multimedia search results cannot consider relationships in them and partially blended. Additionally, the search results' presentation is via linear lists, which cannot support the users' non-linear navigation patterns to explore the multimedia search results. Contrarily, users' are demanding more services from search engines. It includes adequate access to navigate, explore, and discover multimedia information. Our discovery approach allows users to explore and discover multimedia information by semantically aggregating disjoint verticals using sentence embeddings and transforming snippets into conceptually similar multimedia document groups. The proposed aggregation approach retains the relationship in the retrieved multimedia search results. A non-linear graph is instantiated to augment the users' non-linear information navigation and exploration patterns, which leads to discovering new and interesting search results at various aggregated granularity levels. Our method's empirical evaluation results achieve 99% accuracy in the aggregation of disjoint search results at different aggregated search granularity levels. Our approach provides a standard baseline for the exploration of multimedia aggregation search results.

An Architecture for Non-Linear Discovery of Aggregated Multimedia Document Web Search Results

Abdur Rehman Khan[1], Umer Rashid[1], Khalid Saleem[1], and Adeel Ahmed[1]

[1]**Department of Computer Science, Quaid-i-Azam University, Islamabad, 45320, Pakistan**

Corresponding author:
Umer Rashid[1]

Email address: umerrashid@qau.edu.pk

ABSTRACT

The recent proliferation of multimedia information on the web enhances user information need from simple textual lookup to multi-modal exploration activities. The current search engines act as major gateways to access the immense amount of multimedia data. However, access to the multimedia content is provided by aggregating disjoint multimedia search verticals. The aggregation of the multimedia search results cannot consider relationships in them and partially blended. Additionally, the search results' presentation is via linear lists, which cannot support the users' non-linear navigation patterns to explore the multimedia search results. Contrarily, users' are demanding more services from search engines. It includes adequate access to navigate, explore, and discover multimedia information. Our discovery approach allows users to explore and discover multimedia information by semantically aggregating disjoint verticals using sentence embeddings and transforming snippets into conceptually similar multimedia document groups. The proposed aggregation approach retains the relationship in the retrieved multimedia search results. A non-linear graph is instantiated to augment the users' non-linear information navigation and exploration patterns, which leads to discovering new and interesting search results at various aggregated granularity levels. Our method's empirical evaluation results achieve 99% accuracy in the aggregation of disjoint search results at different aggregated search granularity levels. Our approach provides a standard baseline for the exploration of multimedia aggregation search results.

INTRODUCTION

Traditionally, the web contains only the textual content (10). The progressive easy access to the internet has transformed the web into an infinitely complex virtual organism consisting of immense multimedia content (8). The format of the information is now extremely varied. The individual bits of data coming from blogs, articles, web services, picture galleries, etc., are resulting in exponential growth of multimedia data on the web (8; 42). The web is becoming the most ubiquitous platform ever since its birth and has increased in both quantity and quality (56). In 2009, less than 1 petabyte of digital data was created daily (34). It grew to approximately 2.5 exabytes in 2012 and reached 4.4 zettabytes in 2013. On the web, the digital data in different formats created, replicated, and consumed exponentially (40). It is doubling every 2 years. By 2015, digital data grew to 8 zettabytes, and the volume of data will reach 40 zettabytes by the end of 2020 (40).

Keywords-based general web search engines have made early efforts to provide access to multimedia information (35). These search engines required a user to enter one or a few keywords, and the search engines produced the relevant results in a short time (35). Kerne et al. (26) first discussed a new search paradigm called information discovery. They elaborated discovery as a long journey of search that begins with a vague description of a problem, may have an articulated set of criteria during which a searcher specify a query and evaluate the returned information surrogates, and may continue iteratively by re-evaluating the result sets and forming a sense of desired results. Marchioni (39) gives the same idea in a broader perspective by categorizing the search paradigm into an exploratory by incorporating not only lookup searches but learning and investigation activities. Adequate support in the users' search leads to the discovery of new information items.

In contrast, many current search systems assume an exploratory search process as a series of homogeneous steps of submitting a query and consulting search results. Research in information seeking has shown that users go through discrete phases in their search journey, from exploring and identifying preliminary information to refining and narrowing their information needs and search strategies to finalize the search. It is reported as a highly complex problem bridging the different areas of information seeking, interactive information retrieval, and user interface design (21). Moreover, the increasing amount of heterogeneous content on the web has transformed user needs from simple lookup-based queries to broader exploratory queries, requiring the diverse heterogeneous contents to satisfy the desired information needs (42).

Several studies indicate that more intricate tasks resulted in a diversity of the information sought and more varied approaches to information seeking (13). Today, to find interesting multimedia content, an enormous number of users use search engines (17). It has changed users' information need from textual to multi-modal (audio, image, and video) searching. Approximately 40%-50% of users engage in dynamic and unplanned nature of web multimedia searches (58). When the information need is ambiguous and dynamic (e.g., in exploratory search), people often consult more multimedia search results (11). The need for multimedia documents, in this case, increases to 58% (58).

As human information needs and search tasks become complex, the users have to collect and assemble information from diverse information sources. The goal is to compose the most appropriate responses to the tasks at hand in the form of multimedia documents (32). A multimedia document is a collection of co-existing heterogeneous multimedia objects sharing the same semantics (42). Users prefer aggregation of useful multimedia information residing in diverse sources through unified interfaces (32; 42). Similarly, the user interface presenting aggregated contents encouraged participants to view more diversified sources from the search results, and 75% of the participants found this blended approach more comfortable to use (52). The user, click-through rate analysis, reported approximately 33% on augmented multimedia artifacts and nearly 55% multimedia artifacts were found relevant and useful during information exploration activities (53). Overall, users explore the multimedia contents 78% of the time to answer their dynamic information needs (31). Based on recent user behavior in complex information needs, we can easily forecast even more increasing multimedia artifacts consumption from the users in satisfaction of complex information needs and discovering information.

Aggregating disjoint verticals provide access to diverse multimedia documents. A vertical is defined as a specialized assembly of same-typed documents (5). This assembly can be media-specific or domain-specific. The former may include media types (e.g., video, blog, image, etc.). The latter may consist of verticals (e.g., travel, shopping, news, etc.). The aggregation process consists of either Cross-vertical Aggregated Search (cvAS) or Relational Aggregated Search (RAS). The (cvAS) ignores the relation during retrieval and aggregation of multimedia content. The (RAS) considers the relationships in the multimedia information. Despite the key-role of aggregation in bridging the modality gap, this area of research only received limited attention in the past (1). Without substantial creativity, this area of research will soon be abandoned. Our discovery approach aims to bring innovation and creativity in this area of search. We envision bridging the modality gap and shortcomings of current search engines, allowing users to discover multimedia information by aggregating disjoint verticals.

Contributions of our solution have three-folds. Firstly, we presented a creative search results aggregation technique using state-of-the-art semantic analysis. Secondly, we enhanced the current search engine shortcomings in information exploration and discovery activities by augmenting non-linear information seeking patterns. Thirdly, we bridged the information modality gap by encoding the search results in various representations. Our proposed solution is the first to address all of the stated challenges of information aggregation, exploration, and discovery.

The rest of the discussion is organized as follows. We discuss the related work in Section 2. We highlight the deficiencies in the existing approaches and motivation behind this research in section 3. We provide the theoretical foundation and formalization of our proposed approach in Section 4. We present the implementation of the architecture in Section 5. We discuss the experimental results in Section 6. Finally, we compare our approach with state-of-the-art and conclude our discussion in Sections 7 and 8, respectively.

RELATED WORK

Theoretical Background and Frameworks According to Kerne et al., information discovery tasks require finding and collecting relevant information elements; filtering the collected elements; developing an understanding of the found elements and their relationships (27). The overall goal is assembling information and connecting answers to open-ended questions (27). It is a multidisciplinary approach and is built on anomalous states of knowledge (9), berry-picking (7), psychological relevance (23), exploratory search (60), information foraging (41), information seeking (38) and sensemaking (6; 48).

During a task performed by a user, the lack of information triggers the requisite for the information needs. The recognized information needs refer to as an anomalous state of knowledge (9). During the recognized anomalous state of knowledge, the users refer to the information retrieval systems to initiate the information seeking journey (38). During this journey, the user picks relevant information analogous to an organism picking berries in the forest scattered on the bushes; they do not come in bunches. One must select them one at a time (7). Similar to this analogy, the user has to forage for the information and pick items from information patches giving information scent the most (41). Information scents are the cues that help the user in making sense of the provided information. It can be augmented by sensemaking activity. It involves making sense of the data during data analysis, searching for representation, and encoding the data to answer specific task-oriented questions (48). This whole journey can incorporate lookup search, learning, and investigation activities, resulting in a non-linear search pattern.

To do this non-linear search of the information successfully, researchers must leverage their skills and experience to develop search systems that actively engage searchers using semantics, inherent structure, and meaningful categorization (61). In general, the user cannot precisely specify what is needed to resolve recognized information anomaly (9). It often results in the shortcomings of the existing retrieval systems in a scenario where the user cannot correctly formulate information need expression resulted in low precision of the retrieval systems (9). In such case, the users' information needs are not fully satisfied by a single final retrieved set, but by a series of selections of individual bits of information at each stage of the ever-modifying search strategies (7). Hence, these tasks require more recall over precision (55). We must consider the user perspective of information relevance, taking into account how effective the topic of the information retrieved matches the subject of interest and how to represent a piece of information that induces a change in the users' cognitive state (23).

Our proposed solution encodes the multimedia search results semantically by aggregating them in multimedia documents. These documents allow users to pick the most suitable collection of information sufficing their information need the most. Furthermore, we provide multimedia document groups, analogous to patches of information, allowing the user to forage for the information patches giving the most information scent. We increased the information scent for multimedia documents and groups by summarizing the data inside them. Semantically aggregating disjoint multimedia verticals provide the conceptualization of multimedia documents and groups. Furthermore, we augment the non-linear information searching and seeking pattern by instantiating a non-linear graph comprising various granularity levels of search results and proximally similar multimedia content links.

From Federation To Aggregation & Diversification Traditional research mostly centered on assisting users in providing relevant multimedia information federated from various sources and information providers. Kerne et al. (25) provided discovery of search results by dispatching the user query to multiple search engines and extracting the relevant pieces of text snippets and images snapshot on the user interface. Similarly, Sushmita et al. (54) provided a digest-based information exploration approach by collecting various pieces of multimedia information from a variety of sources and encapsulating them in the form of a digest. Afterward, researchers identified the modality gap of information with enormously increasing heterogeneous content on the web, which hindered the information exploration. Hence, the first idea of search results aggregation was presented in a workshop at ACM SIGIR 08 conference (30). Later on, Sushmita et al. (52) advanced this idea towards the blending and evaluation of disjoint multimedia verticals into the web search results (53).

Information aggregation is now widely recognized, considered a bridge that narrows the information modality gap and fosters information exploration. Meanwhile, search engines are also starting to adopt a similar approach in their presentation of the search results (5). The progress in multimedia retrieval presented another challenge in deciding the optimal choice and position of vertical in the search engine result page and was explicitly labeled as a vertical prediction problem. Bakrola et al. (5) provided a solution to this challenge by using implicit feedback of the user in the form of several clicks and then using a support vector machine classifier to predict the most suitable vertical sufficing the given user information needs.

Nowadays, the most common and popular commercial web search engines such as Baidu[1], Bing[2], Google[3], Yahoo![4], Yandex[5] etc, are blending some vertical-specific results, assembled from the other data sources into the linear ranked list of standard results. Moreover, a recent trend focuses on the information diversification aspects of the information (57). It usually involves integrating more diverse verticals (e.g., other than image, news, video, and web). This diversification may include the integration of verticals from social media, shopping, movies & dramas, maps, songs, etc. However, this integration of the verticals is mostly partial-blended (42). The relationship between the multimedia artifacts inside each disjoint vertical is often ignored.

The aggregation of the multimedia artifacts demands a better solution to enhance user interaction with the search results. It is essentially a very broad problem and answered by (RAS) techniques. The researchers leveraged some effort in (42), they performed (RAS) using textual, visual, and acoustic descriptors of the multimedia contents. However, this aggregation was provided using a generic similarity measure for each modality and ignored the semantics relationships in aggregated multimedia documents. In (2), researchers presented a stacked auto-encoders model for aggregation of the disjoint verticals. However, their research addresses a small aspect of the aggregation and ignores information exploration perspectives.

Renovation in Information Exploration Data-Models & Semantic Web The current practices for information exploration include presenting the aggregated verticals as a linear list (55). It is due to a lack of data-model flexibility. Initially, using the semantic web techniques and ontologies was perceived as a promising start. For instance, Tablan et al. (55) presented an open-source semantic framework providing indexes and searches using document structure, metadata, annotations, and semantics through linked open data. The architecture supported both; information seeking and exploration & discovery tasks by two distinct user interfaces designed, respectively.

Similarly, Lisena et al. (37) developed a modern web application for music exploration and discovery using semantic RDF graphs to establish links between entities and relationships among them. Khalil et al. (28) used inference techniques on the semantic linked open data to produce notably unique information fostering discovery. However, due to scalability challenges in exploiting the whole web of Linked Data limits the practicality of this aspect (19).

Similarly, in (24), researchers provided semantic data representation in a hyperbolic tree format. Their framework consists of a 3-layers hyperbolic tree-based modal approach that takes the input in the form of keywords from the user. The information is then presented in the form of a graph. The 3-layer approach divides the complexity of information in each layer. It reduces the confusion caused by information overload and enhances significant interaction and navigation. Similar to our proposed approach, their graph data-model provides highlighting, node describing, zooming, panning, and linking functionalities.

More researchers are presently making an effort to provide a generalized approach to exploring and discovering multimedia artifacts on the web. It includes mixing different aspects of data-model, diverse information aggregation, and visualization. For instance, in (42), researchers provide a generalized framework for relational aggregation of the multimedia artifacts belonging to disjoint sets using a graph-based visualization and exploration of a multimedia search result space. Similarly, in (62), researchers developed a discovery engine for artificial intelligence research. Their architecture crawls the web, downloads the research papers from various journal websites, and performs full-text indexing using a cosine similarity measure. It builds a similarity-based network having similarity links in documents. Users' stars, clicks, and tweets are primarily used to reinforce the graph's essential connections.

However, the past approaches focus on using a domain-specific dataset and data-model using generic textual and visual similarity metrics. We establish a data-model using the semantics that exists inside the data. Specifically, we semantically found part-of or containment relationships in the multimedia artifacts. Moreover, we also instantiate similarity links among the multimedia artifacts that allow navigation to

[1] http://www.baidu.com/
[2] https://www.bing.com/
[3] https://www.google.com/
[4] https://www.yahoo.com/
[5] https://yandex.com/

similar multimedia artifacts. We opt to keep the data-model as generic as possible without relying on domain knowledge and explicit feedback, making our solution implementable on a wide range of domains.

PROBLEM & MOTIVATION

The users' complex information-seeking behavior is modeled as a non-linear journey requiring adequate support during the navigation of the information space (45). Users' forage for the information (7). Their complex information needs are not sufficed through the current ideology of returning the most precise information in response to the given queries (49). Instead, users picking the most interesting items like barries from various patches of information, providing more information scent ratio to the effort required for examining the information. It results in a non-linear information searching pattern of users (49).

The search engines are more tuned towards simple lookup searches favoring precision over recall (55). However, even though they have recognized the users' multimedia information needs and started to blend some vertical-specific results assembled from the other data sources (5). The current practices of presenting information in a linear ranked list of standard results limit information exploration (42). Furthermore, the integration of the verticals is mostly partial-blended (42), which may suffice in simple lookup searches when a user knows what to look for; however, this strategy inadequately support complex information exploration and discovery tasks (55). These tasks go beyond simple keyword-based queries. Users often have difficulties in information need expression, and they usually are dynamic (44). Such tasks require more recall over precision and diversity of information sources (55). It challenges the current practices of displaying the search results belonging to different verticals as disjoint sets (42).

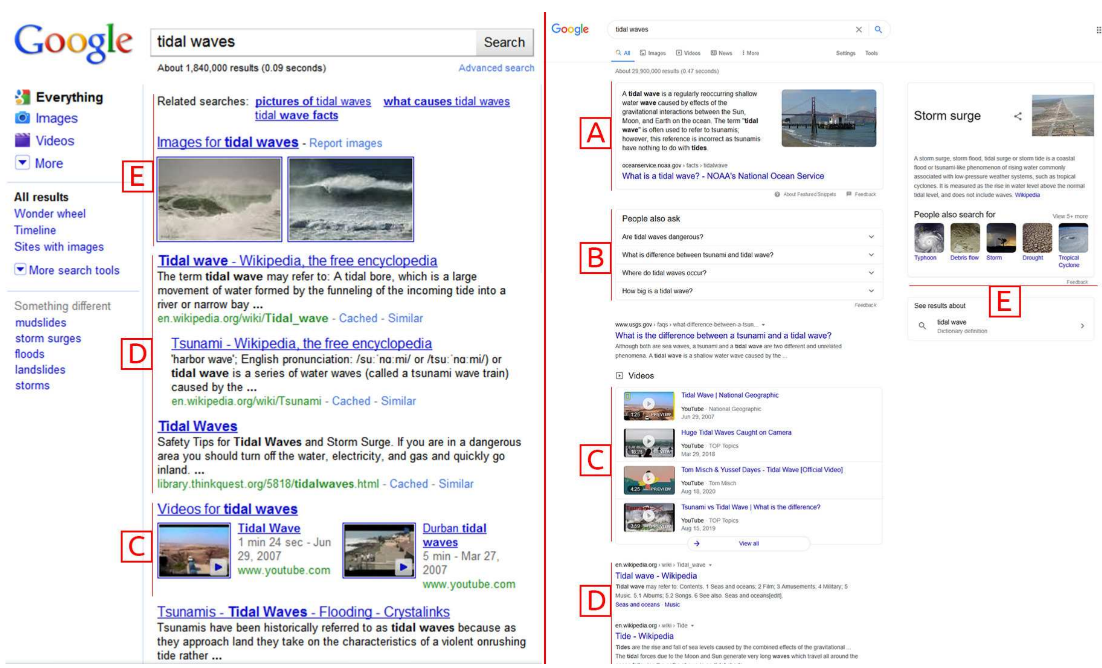

Figure 1. Comparison of Google SERP between 2010 (left) and 2020 (right). (A) Enhanced Snippet, (B) Question Answer Vertical, (C) Videos Vertical, (D) Web Vertical, (E) Images and Related Searchers

On the other hand, the search engines remain almost the same as they were about a decade ago. There exist numerous problems (**P**) with current search engines. The figure 1 shows the difference between the Google Search Engine Results Page (SERP) back in 2010 (51) and now in 2020 (22). The verticals are integrated as disjoint components (P_1). The relationships between multimedia objects are ignored (P_2). The information presented is still displayed as linear lists (P_3). This presentation of the general search engines' information may suffice for simple lookup tasks but lacks adequacy for complex exploratory and discovery tasks (29). These tasks require increased recall over precision (P_4), information scent (P_5), and sensemaking (P_6). The existing exploration approaches' deficiencies demand a better mechanism to encode and present the multimedia information for discovery (P_7).

ARCHITECTURE DESIGN: DEFINITION, FORMALIZATION & INSTANTIATION

Existing techniques are usually specific to a problem and employed on a particular dataset. Many researchers consider one side of the discovery, such as information diversification, visualization, data-modal etc., and ignore the other factors highlighted in the previous section. To the best of our knowledge, a generalized multimedia search results discovery mechanism, particularly in aggregated search, is the first to address in this research. Notably, we provided a balanced architectural approach for information discovery, emphasizing the dataset, data-model, information diversification equally. We used real-dataset retrieved from the search engines in real-time. We instantiated a non-linear graph data modal consisting of diverse information while preserving the semantics and similarity relationships. Finally, we provided a theoretical background to foster exploration and discovery activities.

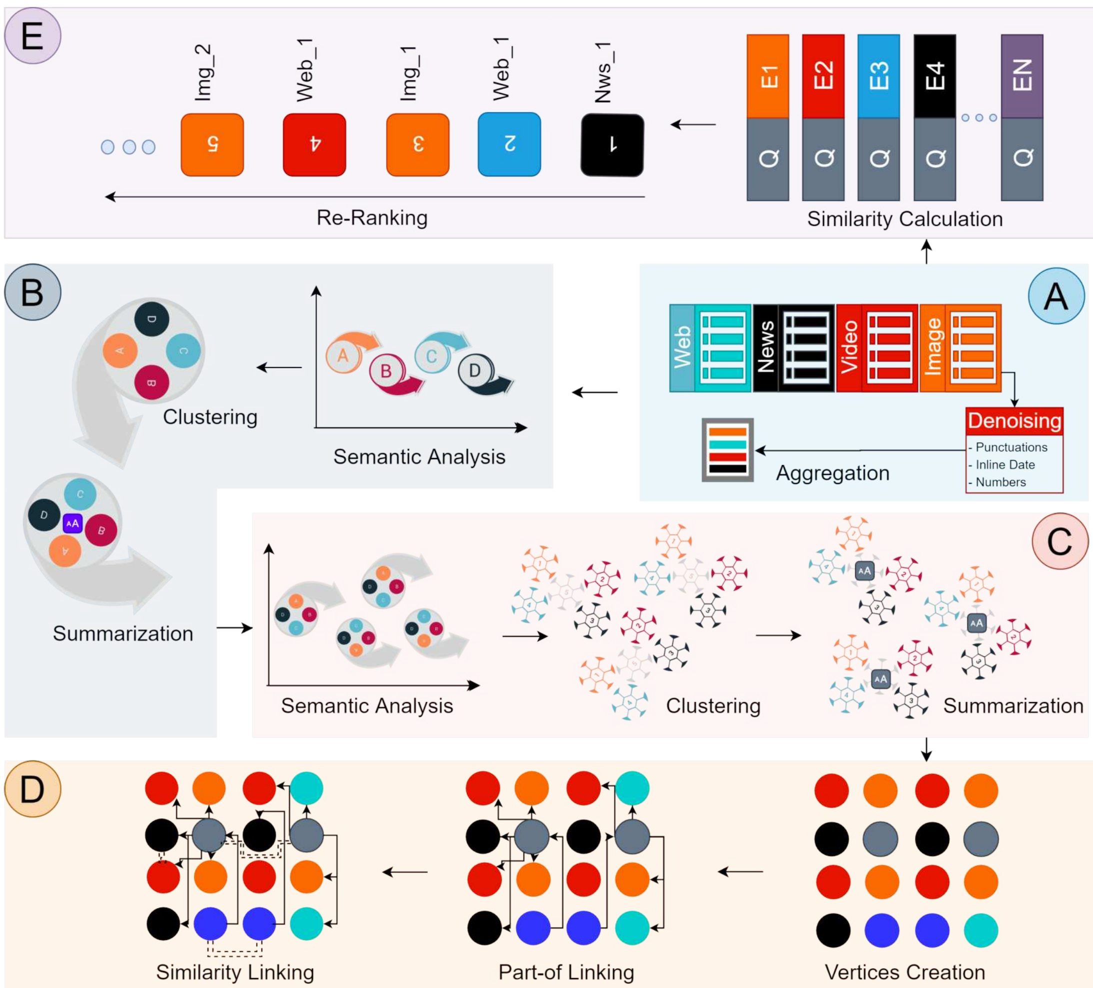

Figure 2. Discovery Architecture Design. Component (A) Search Results Aggregation, (B) Multimedia Documents Creation, (C) Multimedia Documents Grouping, (D) Graph Instantiation, (E) Semantic Lookup List

Since you are using A, B, C in the image,
I would use the same symbols in the description
in place of (i), (ii), etc

Our component-based architecture design includes sub-components. Each sub-component produces a consumable output. There are five main components, referred to (i) Search Results Aggregation; (ii) Multimedia Document Creation; (iii) Multimedia Documents Grouping; (iv) Graph Instantiation; (v) Semantic Lookup List, components as illustrated in figure 2. Each component is concerned with delegated responsibility, and their internal working is separate from each other. A discussion on each component is provided in the following sections.

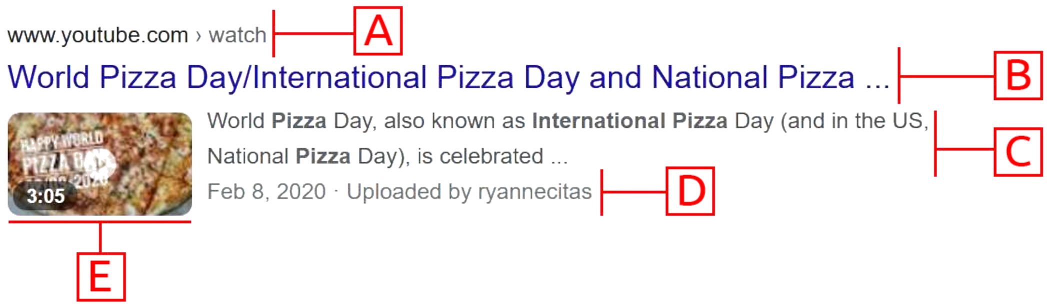

Figure 3. Anatomy of search result snippet. (A) URI, (B) Title, (C) Description, (D) Date, (E) Thumbnail

Search Results Retrieval & Aggregation

Search results retrieved from the search engines are presented in the form of disjoint verticals. Adversely, users' information needs are becoming complex and multi-modal, requiring the employment of multimedia artifacts for satisfaction. To aggregate scattered disjoint verticals (P_1), we introduced a search results aggregation component. The aggregation process of this component is subdivided into three steps.

Vertical Retrieval

Definition: We define vertical as a specialized assembly of same-typed search results and retrieval as a process of obtaining them from some external source.

Formalization: Let S be the set of α, β, γ, λ respectively given as $S = \{\lambda, \alpha, \beta, \gamma\}$, where α is defined as a set of video snippets V given as $\alpha = \{\{V_1^\Psi, V_2^\Psi, V_3^\Psi, ..., V_n^\Psi\}, \{V_1^\Phi, V_2^\Phi, V_3^\Phi, ..., V_n^\Phi\}\}$, β is defined as a set of news snippets N given as $\beta = \{\{N_1^\Psi, N_2^\Psi, N_3^\Psi, ..., N_n^\Psi\}, \{N_1^\Phi, N_2^\Phi, N_3^\Phi, ..., N_n^\Phi\}\}$, γ is defined as a set of image snippets I given as $\gamma = \{\{I_1^\Psi, I_2^\Psi, I_3^\Psi, ..., I_n^\Psi\}, \{I_1^\Phi, I_2^\Phi, I_3^\Phi, ..., I_n^\Phi\}\}$, and λ is defined as a set of web snippets W given as $\lambda = \{W_1^\Psi, W_2^\Psi, W_3^\Psi, ..., W_n^\Psi\}$, where ψ and Φ denotes the textual and visual modality associated with a snippet respectively.

Instantiation: We retrieve top hundred search results from each **web**, **news**, **image**, and **video** verticals. Since exploratory and discovery tasks require increased recall over precision (P_4), we chose to retrieve maximum search results from the API provider. With each search result, we preserve the metadata associate with it, including title, description, date, URL, and thumbnail (where available). The verticals are retrieved from the Google search engine in real-time because Google is highly preferred by web users (3). The table 1 shows the vertical retrieval parameters. Figure 3 outlines the title, description, date, thumbnail, and URI of each snippet.

Table 1. Parameters for verticals retrieval

Vertical	# of results (n)	Source	Modality	Feature(s)
Web	≤ 100	Google[3]	Textual	Title, Description, URL
Video	≤ 100	Google[3]	Textual + Visual	Title, Description, URL, Thumbnail, Date
News	≤ 100	Google[3]	Textual + Visual	Title, Description, URL, Thumbnail, Date
Image	≤ 100	Google[3]	Textual + Visual	Title, URL, Thumbnail

Verticals Aggregation

Definition: We define verticals aggregation as a single container of all the retrieved disjoint verticals.

Formalization: Let X be the subset of S consisting of the all the elements λ, α, β and γ from S. We consider only textual modality information given as $X = \sum_{i=1}^{n} S_i^{\psi}$.

Instantiation: Each retrieved snippet has unwanted data, including HTML tags, numbers, special characters, inline date, etc. These impurities add meaning to neither the semantics nor similarity analysis. We are restraining to perform extra pre-processing steps such as stopwords removal and stemming. It results in the loss of contextual information necessary for semantic analysis. Afterward, we preserve the scattered disjoint verticals textual data inside a single container as a linear list.

Multimedia Document Creation

Previous studies indicate user interest in exploring multimedia documents encapsulating relevant multimedia objects during information exploration (42). We define a multimedia document as a semantic container of similar content belonging to multiple modalities. Instead of providing a linear list of snippets, which forces web users to locate scattered relevant multimedia objects from disjoint verticals, we give document-based multimedia exploration (P_6). The multimedia document semantically gathers the scattered multimedia objects belonging to various disjoint verticals. This process is again sub-divided into three steps.

Semantic Analysis

Definition: We define semantic analysis as a process of obtaining semantic information (relatedness and containment) from transformed multidimensional vector representation of search results textual data (P_2).

Formalization: $\forall\, x\, \varepsilon\, X$ let E_x be the set of sentence embedding given as $E_x = \{e_1, e_2, e_3, ..., e_n\}$, Where each element in E_x is represented by $e = \{r_1, r_2, r_3, ..., r_{768}\}$ and $r\varepsilon(\mathbb{R})$.

Instantiation: Firstly, we transformed each multimedia snippet in the aggregated list into sentence embeddings. This transforms each snippet into a multidimensional vector space for semantic analysis. Since each snippet contains minimal textual description, sentence embedding is deemed a better choice over the Doc2Vec technique.

Clustering

Definition: We define clustering as a process of grouping the search results, having highly related intra-group coherence and otherwise for the inter-group search results.

Formalization: $E_x = c_1 \cup ... c_i \cup c_n; c_i \cap c_j = \emptyset (i \neq j)$, where E_x denotes original data, c_i, c_j are clusters of E_x and n is the number of clusters. Let C be the set of clusters of E_x given as $C = \{c_1, c_2, c_3, ..., c_n\}$, where each cluster contains a set of coherent text t and $c = \{t_1, t_2, t_3, ..., t_n\}; t\varepsilon C$.

Instantiation: We performed the agglomerative clustering on all the semantic search results vectors. Agglomerative clustering is chosen for due to flexibility in the clustering process as it allows the clusters to be obtained using cut-off criteria instead of predefined number of clusters. This process groups similar search results in various buckets, called multimedia document.

Summarization

Definition: We define summarization as a process of extracting the most representative words from the bucket of search results.

Formalization: Let K be a set of words sequence k from $c\ \varepsilon\ C$, generated by text summarizer representing the collection of text given in c as $K = \{k_1, k_2, k_3, ..., k_n\}$, let M_d be the multimedia document, we formed M_d by mapping function $M_d = \forall(C)\forall(K)(f(C) = f(K) \rightarrow C = K)$, where $\forall x \varepsilon X, \exists \kappa \varepsilon K, f(K) = X$.

Instantiation: To enhance sensemaking (P_6), instead of merely labeling a multimedia document by assigning predefined categories, we are performing summarization based on the text of the snippets inside the multimedia document. Specifically, we perform extractive text summarization techniques to extract the combination of the most representing text inside the multimedia document for its representation.

Multimedia Document Grouping

Prior research has shown that web user information exploration behavior is analogous to a foraging animal in the forest (41). They look for the patches containing more information scent as compared to the effort performed. In traditional linear list presentation of the search results, a user has extreme difficulty locating the appropriate patches of information and comprehending search results space (47). In this component, we grouped multimedia documents to provide patches of information and enhance search results in space comprehension (P_5). This process is sub-divided into three steps.

Semantic Analysis

Definition: We define semantic analysis as a process of obtaining semantic information (relatedness and containment) from transformed multidimensional vector representation of multimedia documents.

Formalization: We take the K which is the set consisting of text summarized from each multimedia document. $Y = \sum_{i=1}^{n} K_i$, we produce the set Y from K to perform semantic analysis. $\forall\ y\ \varepsilon\ Y$ let M_x be the set of sentence embedding given as $M_x = \{e_1, e_2, e_3, ..., e_n\}$, where each element in M_x is represented by $e = \{r_1, r_2, r_3, ..., r_{768}\}; r\varepsilon(\mathbb{R})$.

Instantiation: Firstly, we extract summaries of multimedia documents and aggregated them inside a linear list. Then we performed semantic analysis on each multimedia document summary using sentence embeddings. This transformed each multimedia document to a multidimensional vector space for semantic analysis. Similarly, since each multimedia document contains minimal textual description, sentence embedding is deemed a better choice over the Doc2Vec technique.

Clustering

Definition: We define clustering as a process of grouping the multimedia documents having high intra-group relatedness and otherwise for the inter-group multimedia documents.

Formalization: Let $M_x = c_1 \cup ... c_i \cup c_n; c_i \cap c_j = \emptyset (i \neq j)$, where M_x denotes original data, c_i, c_j are clusters of M_x and n is the number of clusters.

Instantiation: We performed the agglomerative clustering on all the semantic vectors of multimedia document summaries. Similarly, agglomerative clustering is chosen for provided flexibility creation process of clusters using a cut-off criteria. This process groups similar multimedia documents in various buckets.

Summarization

Definition: We define summarization as a process of extracting the most representative words from the bucket of multimedia documents.

Formalization: Let $\gimel$ be the set of clusters of M_x given as $\gimel = \{c_1, c_2, c_3, ..., c_n\}$, where each cluster contains a set of similar text t $c = \{t_1, t_2, t_3, ..., t_n\}; t\varepsilon\gimel$. Text summarizer $\Re$ produces a set of words sequence j from $c\ \varepsilon\ \gimel$ representing the collection of text given in c as $\Re = \{j_1, j_2, j_3, ..., j_n\}$. Similarly, Let M_c be the multimedia document cluster, we formed M_c by mapping function $M_c = \forall(\gimel)\forall(\Re)(f(\gimel) = f(\Re) \rightarrow \gimel = \Re)$, where $\forall y\varepsilon Y, \exists j\varepsilon\Re, f(\Re) = Y$.

Instantiation: We call each generated bucket of multimedia documents from the clustering process a multimedia document group. To enhance sensemaking, instead of merely labeling a multimedia document group by arranging them in taxonomic order, we perform summarization based on the multimedia document summary. The summarization process is performed using extractive text summarization technique. This extracts the most representing text inside the multimedia document group.

Graph Instantiation

Present search engines display the search results in a linear list and often ignores the relationship between multimedia content. As a result, users have to navigate the results space and berry-pick the relevant items of interest (7). This results in a non-linear searching pattern of a user in the exploration of information (7). To overcome these challenges, we instantiated a non-linear graph augmenting the users' non-linear exploratory information-seeking behavior while preserving the relationships (P_2). This process is sub-divided into three steps, as well.

Vertices Creation

Definition: We define vertex as an atomic data structure encapsulating the complete details of the representing entity.

Formalization: Let a graph G be a set of vertices V and edges E, given as $G = (V,E)$ and vertices V represent all the vertical snippets, multimedia documents and clusters given as $V = \{S, M_d, M_c\}$.

Instantiation: Firstly, we represented each multimedia document group, multimedia document, and multimedia snippet as a vertex. We associate with each vertex the metadata. It includes a text summary for the multimedia documents and multimedia documents. Similarly, metadata belonging to the multimedia snippet include their title, description, URI, date, and thumbnail (where available).

Part-of Linking

Definition: We define part-of linking as a process of establishing containment relationship between vertices.

Formalization: The edge (δ) between the S and M_d denotes the part-of relationship given as $\delta : \forall x \varepsilon S, \exists d \varepsilon M_d, f(M_d) = S$. Similarly, edge ($\delta$) between the M_d and M_c denotes the part-of relationship given as $\delta : \forall m \varepsilon M_d, \exists c \varepsilon M_c, f(M_c) = M_d$.

Instantiation: Since a multimedia document is a part of some multimedia documents group, similarly, a multimedia snippet is a part of some multimedia document, the edges established between them represents the part-of (or containment) relationship.

Similarity Linking

Definition: We define similarity linking a process of establishing proximally similarity-based relationship between vertices.

Formalization: Edges (δ) among M_c denotes the similarity relationship based on Cartesian product of M_c given as

$$\delta : \begin{cases} M_c \times M_c = \sum_{i=1}^{n} \sum_{j=i+1}^{n} J(M_i^c, M_j^c), & if J > \theta \\ \emptyset, & \text{otherwise} \end{cases}$$

Similarly, edges (δ) among M_d in each M_c denotes the similarity relationship based on Cartesian product of M_d within M_c given as

$$\delta : \begin{cases} \sum_{k=1}^{n} M_k^c \forall M_d \varepsilon M_k^c : M_d \times M_d = \sum_{i=1}^{n} \sum_{j=i+1}^{n} J(M_i^d, M_j^d), & if J > \theta \\ \emptyset, & \text{otherwise} \end{cases}$$

Where J is the Jaccard similarity defined as $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$ and θ is the average similarity score of all the selected vertices pairs in the graph.

Instantiation: Exploratory search also involves navigation of proximally similar multimedia documents in the collection (50). It helps a user explore the environment to understand better how to exploit it, selectively seek and implicitly obtain cues about coming steps (50). Hence, we provide navigational links to proximally similar multimedia document groups and multimedia documents. These links are established on the Cartesian pairs of multimedia documents groups if there is a high proximal similarity between the source and destination vertices. The same procedure is performed for multimedia documents inside each multimedia document group. We chose the Jaccard similarity measure because it is computationally less expensive than other similarity techniques (42).

Semantic Lookup List

At present, the aggregation of the verticals on the major search engines is provided as partially-blended. The relationship between the multimedia snippets in those retrieved disjoint verticals is ignored (42). On the other hand, information lookup is an eminent component of information exploration and discovery, and linear lookup lists have proven to be effective in information lookup (55). To overcome this challenge of disjoint verticals relation-less aggregation of the verticals and provide ease in lookup searches, we introduce a semantic lookup list component that fully-blends the disjoint verticals semantics of the multimedia snippets (P_1). This component is divided into two sub-components.

Similarity Calculation

Definition: We define similarity calculation a process of extracting numeric similarity score between pairs of text using a textual similarity measure.

Formalization: Let the E_x be the same set of sentence described previously, we also transformed user query Q as a sentence embedding Q_x represented by $Q_x = \{r_1, r_2, r_3, ..., r_{768}\} : r\varepsilon(\mathbb{R})$.

Instantiation: In this part, we transform the user query itself into the sentence embeddings. This transformation eliminates the data representation gap.

Re-Ranking

Definition: We define re-ranking as a process of arranging search results in descending order of query and search results embedding pairwise intra-similarity scores.

Formalization: We define $SIM(Q_x, e)$ as cosine similarity function, calculating pairwise vectors similarity of Q_x and e, given as $SIM(Q,e) = \frac{Q_x.e}{\|Q\| \times \|e\|}$, where $L_s = \sum_{i=1}^{N} 0 \leq SIM(Q, e_i) \leq 1$. Using similarity scores L_s, we define L_r the ranked linear search results list, sorted in descending order of similarity, given as $L_r = \{l_i \leq l_{i+1} \leq l_{i+2} \leq ... \leq l_N\}$, where $l\varepsilon S$ & $N = |S|$.

Instantiation: In lookup searches, the ordering of information is mandatory. The most relevant information must be present on the most top. The search engines return disjoint ranked verticals. To calculate the ranking order for snippets belonging to aggregated disjoint verticals, we perform a re-ranking operation on the pair-wise (query and each snippet embedding) obtained semantics using a cosine similarity measure. We use cosine similarity because our query and search results are in vector representation. We re-rank each multimedia snippet in their descending order of similarity, allowing the most relevant snippet to appear first on the linear list.

ARCHITECTURAL IMPLEMENTATION

We implemented our architecture in *Python*3 programming language using publicly available libraries. Search results are retrieved using freely available APIs to fetch the verticals from a search engine in real-time. We used Google[3] search engine to retrieve the search results belonging to the web, news, image, and video verticals. We preserved the metadata associated with each snippet, such as the URL, title, date, length, description, and thumbnail, where available. For text summarization, we used *LexRank*[6] extractive text summarization algorithm. Semantic analysis is done using $SBERT's$[7] sentence embedding on pre-optimized $bert-base-nli-mean-tokens$[8] pre-trained modal and agglomerative clustering using the ward's linkage method from the *sklearn*[9] python library to obtain the clusters. We used *Networkx*[10] python library to instantiate an undirected network to build the graph. Each node represented either a web snippet, multimedia document, or a multimedia document cluster. The snippet nodes attribute includes their metadata. The multimedia document and multimedia document cluster nodes attribute include their summarized text. Figure 4 shows the visualization of the instantiated graph generated from *Cytoscape* [11].

[6] https://gist.github.com/rodricios/fee45381356c8fb36004
[7] https://pypi.org/project/sentence-transformers/
[8] https://github.com/UKPLab/sentence-transformers/blob/master/docs/pretrained-models/nli-models.md
[9] https://scikit-learn.org/
[10] https://networkx.github.io/
[11] https://cytoscape.org/

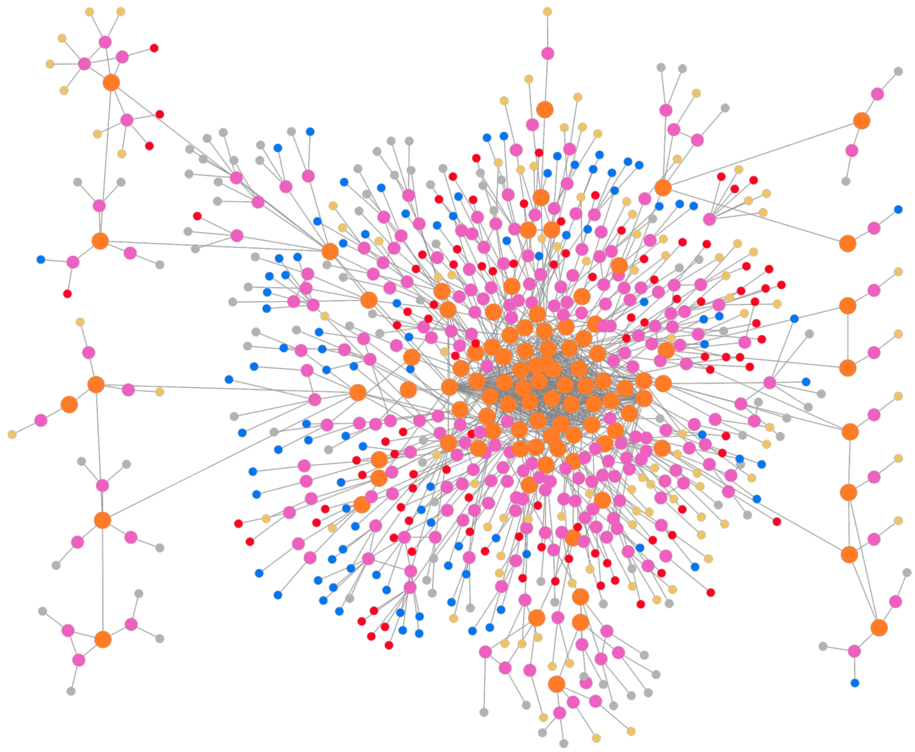

Figure 4. Visualization of the instantiated graph, the orange and pink color represents cluster and multimedia document respectively. The rest denotes snippets belonging to disjoint verticals.

RESULTS

There is still no standard empirical evaluation measures for evaluating the aggregated search approach effectiveness (36). These approaches are mostly considered in terms of the achieved precision & recall (42) and judgment reports from the human experts (46). Calculating precision and recall in our case is a non-trivial task. It is mainly due to the nature of the data. Therefore, we used a real dataset with no prior labeling by human experts. Our empirical evaluation measures mostly depend on metrics requiring no initial labeling of data. We used internal clustering stability measures to evaluate the internal cluster model stability (59), and clustering accuracy based on the judgment of the human experts (45). We obtained accuracy and stability scores by dispatching pre-defined queries on Google's real dataset.

We collected queries from the recently published ORCAS (14) dataset consisting of 10 million distinct records. Selecting all queries in the dataset for evaluation purposes was not practical. Hence, we performed bi-gram and tri-gram query analysis on the ORCAS dataset. Afterward, we selected 25 queries from the top 100 most repeating bi-gram and tri-gram combinations. The average query length for this evaluation was set to 2.5 words. The chosen length was due to a recent study in (16) indicating average user query length between 2.44 and 2.67 words, which confirms that users' information needs are becoming exploratory. Since in exploratory search, user needs are ambiguous, and the primary objective is to gain an overview of the information. Users type short queries instead of well-articulated longer queries as in the lookup search scenarios (4). We selected queries covering broad aspects. Therefore, an average query length of 2.5 words was considered based on the average of 2.44 and 2.67 words.

Internal Clustering Parameterizing

We used agglomerative clustering for the creation of multimedia documents and multimedia documents groups. We specified cut-off threshold criteria for the cluster creation process θ to form the desired number of clusters. We chose θ empirically by determining the best possible average mean value of internal cluster stability measures. We used a well-known Silhouette Coefficient (43), Calinski-Harabasz (12) and Davies-Bouldin (Davies) index to calculate internal cluster stability. We calculated the mean average

Table 2. Empirical multimedia documents grouping via clustering results

Experiment	Iteration	Optimal θ_2	# of Clusters	Silhouette Coefficient	Calinski-Harabasz index	Davies-Bouldin Index
1	1	17	79	0.08	4.97	1.79
	2	19	5	0.05	3.15	2.00
	3	14	90	0.05	2.28	1.06
	4	17	54	0.05	2.66	1.47
	5	13	54	0.07	3.32	1.08
	Mean	16	56.4	0.06	3.28	1.48
2	1	17	76	0.04	2.83	1.81
	2	16	81	0.04	2.61	1.42
	3	15	113	0.05	2.51	1.21
	4	15	93	0.05	2.54	1.24
	5	16	97	0.06	2.73	1.28
	Mean	15.8	92	0.04	2.65	1.39
3	1	15	61	0.05	2.93	1.58
	2	16	106	0.06	2.57	1.33
	3	15	69	0.05	2.84	1.27
	4	15	77	0.04	2.69	1.24
	5	16	103	0.05	2.54	1.29
	Mean	15.4	83.2	0.05	2.71	1.34
4	1	16	82	0.06	2.63	1.39
	2	16	55	0.06	2.86	1.56
	3	16	85	0.04	2.64	1.43
	4	15	114	0.04	2.30	1.16
	5	15	78	0.05	2.58	1.18
	Mean	15.6	82.8	0.05	2.60	1.34
5	1	15	78	0.04	2.76	1.27
	2	16	50	0.05	3.01	1.62
	3	15	58	0.05	2.78	1.40
	4	14	60	0.06	3.01	1.13
	5	14	64	0.05	2.74	1.14
	Mean	14.8	62	0.05	2.86	1.31
Mean Average		15.52	75.28	0.05	2.82	1.37

value of θ_1 by performing five experiments and taking their mean value to create multimedia documents, as displayed in the table 2. Based on the obtained θ_1 threshold, we again repeated the same procedure for multimedia documents clustering as displayed in table 3 to obtain $_2$. Finally, we parameterized the clustering model for multimedia documents and multimedia documents clusters based on empirically obtained values, as displayed in table 4 and table 5 respectively.

Clustering Precision

Precision is referred to as a fraction of relevant retrieved out of total relevant results (42). In clustering, precision is a fraction of relevant results out of total results inside a cluster. Precision is mostly calculated by cross-matching obtained cluster results with correct labeled data. In a real dataset, the labeling of data is unavailable. We logged the search results retrieved from the pre-defined queries during the empirical internal clustering model parameterization process to overcome this challenge. These logged search results were then presented to two human experts to label relevant and irrelevant search results inside each cluster. The first human expert is a graduate in education and had no prior knowledge about computing-related technical aspects. The second human expert is a graduate in computer science and had substantial knowledge about computing technical aspects, including the concept of clustering. This diversity in the background helps in obtaining unbiased validation of our clustering approach.

Table 6 show the results obtained from the human experts. We run a total of 25 experiments, divided into 5 iterations. From each iteration, we obtained the mean results. Afterward, we took the mean average of 5 iterations. This process was repeated for both; the multimedia documents and multimedia documents groups. The results show no significant change in the relevancy judgment scores from both the novice judge (99.53%) and the expert judge (99.13%) for multimedia documents. Similar results were achieved for multimedia documents groups from the novice judge (99.80%) and expert judge (99.61%). There is a moderate amount of agreement ($\kappa = 0.474$) between the novice and expert judges for multimedia documents. Similarly, there is a fair amount of agreement ($\kappa = 0.398$) between the novice and expert judges for multimedia documents.

Table 3. Empirical multimedia documents creation via clustering results

Experiment	Iteration	Optimal θ_1	# of Clusters	Silhouette Coefficient	Calinski-Harabasz index	Davies-Bouldin Index
1	1	15	174	0.11	3.40	1.15
	2	9	200	0.15	5.88	0.88
	3	12	240	0.12	3.82	0.83
	4	14	148	0.11	3.59	0.98
	5	10	248	0.15	5.60	0.75
	Mean	12	202	0.13	4.46	0.92
2	1	14	209	0.10	3.39	0.97
	2	15	157	0.09	3.56	1.27
	3	15	160	0.11	3.36	1.28
	4	16	123	0.08	3.62	1.50
	5	13	255	0.15	4.08	0.74
	Mean	14.6	180.8	0.11	3.60	1.15
3	1	13	187	0.08	3.49	1.08
	2	15	180	0.10	3.36	1.11
	3	13	191	0.10	4.01	1.01
	4	16	102	0.09	4.51	1.59
	5	15	178	0.12	3.48	1.12
	Mean	14.4	167.6	0.10	3.77	1.18
4	1	14	190	0.12	3.94	1.01
	2	14	160	0.10	3.80	1.20
	3	14	205	0.09	3.06	1.06
	4	13	249	0.12	3.22	0.81
	5	13	186	0.17	4.47	0.88
	Mean	13.6	198	0.12	3.70	0.99
5	1	14	166	0.10	4.04	1.16
	2	13	192	0.09	3.85	1.00
	3	13	182	0.10	3.58	1.09
	4	12	204	0.10	4.06	0.96
	5	12	192	0.14	5.46	0.92
	Mean	12.8	187.2	0.11	4.20	1.03
Mean Average		13.48	187.12	0.11	3.95	1.06

Table 4. Multimedia documents clustering model parameters

Parameters	Description	Value
n_clusters	# of clusters to find	None
affinity	Metric to compute linkage	Euclidean
distance_threshold	The linkage distance threshold for merging clusters	13.48
linkage	Distance method between set of observations	Ward

Table 5. Multimedia documents grouping clustering model parameters

Parameters	Description	Value
n_clusters	# of clusters to find	None
affinity	Metric to compute linkage	Euclidean
distance_threshold	The linkage distance threshold for merging clusters	15.52
linkage	Distance method between set of observations	Ward

Table 6. Clustering precision. Clusters precision for (1) Multimedia documents, (2) Multimedia documents groups, Relevancy scores by (a) Novice judge (b) Expert judge

Experiment	Iteration	Precision %age (1a)	Precision %age (1b)	Precision %age (2a)	Precision %age (2b)
1	1	100.00	100.00	100.00	92.00
	2	100.00	100.00	100.00	100.00
	3	100.00	96.70	100.00	100.00
	4	96.60	88.00	100.00	100.00
	5	100.00	100.00	100.00	100.00
	Mean	99.32	96.94	100.00	98.40
2	1	100.00	100.00	100.00	100.00
	2	100.00	100.00	100.00	100.00
	3	100.00	100.00	100.00	100.00
	4	100.00	100.00	100.00	100.00
	5	100.00	100.00	100.00	100.00
	Mean	100.00	100.00	100.00	100.00
3	1	100.00	100.00	100.00	100.00
	2	100.00	100.00	100.00	100.00
	3	100.00	100.00	100.00	100.00
	4	100.00	100.00	100.00	100.00
	5	96.60	100.00	100.00	98.20
	Mean	99.32	100.00	100.00	99.64
4	1	98.10	100.00	98.20	100.00
	2	100.00	100.00	96.90	100.00
	3	96.90	96.90	100.00	100.00
	4	100.00	100.00	100.00	100.00
	5	100.00	100.00	100.00	100.00
	Mean	99.00	99.38	99.02	100.00
5	1	100.00	100.00	100.00	100.00
	2	100.00	100.00	100.00	100.00
	3	100.00	96.60	100.00	100.00
	4	100.00	100.00	100.00	100.00
	5	100.00	100.00	100.00	100.00
	Mean	100.00	99.32	100.00	100.00
Mean Average		99.53	99.13	99.80	99.61
Average		99.33		99.71	
Cohen's Kappa		0.474		0.398	

Table 7. Comparison of the proposed approach with state-of-the-art

Category	Parameter	Values	(25)	(44)	(55)	(18)	(33)	(28)	(37)	(42)	((57)	(62)	(24)	Proposed Approach
			State-of-the-Art											
Searching	Search Type	Full text	+	+	+	+	+	+		+	+	+		+
		Fielded			+			+	+					
		Semantic			+			+	+				+	+
		Federated	+		+		+			+	+			+
	Search Results Granularity	Snippets	+				+	+	+		+		+	+
		Document		+	+	+						+		+
		Document Clusters												+
	Search Activity	Lookup							+			+		+
		Exploratory	+	+	+	+	+	+		+	+	+	+	+
		Discovery	+	+	+		+	+	+		+		+	+
Data	Information Source	Web	+				+				+	+		+
		Repository	+	+	+	+	+	+	+	+	+		+	
		Real	+				+				+			+
	Data Model	Linear		+	+	+	+				+			+
		Non-linear	+		+			+	+	+		+	+	+
	Data Relation	Part-of			+			+	+	+			+	+
		Similarity	+	+	+	+	+	+	+	+	+	+		+
		Semantic			+		+	+						+
Information Retrieval	Media Source	Textual	+	+	+	+	+	+		+	+	+	+	+
		Multimedia	+						+	+	+			+
	Retrieval Modal	Monomodal		+	+	+	+	+				+	+	
		Cross-Modal	+						+	+	+			+

COMPARISON & DISCUSSION

Our approach outperforms in terms of accuracy (99%) in comparison to the approach provided by Achsas et al. (89%) (2). It mainly can be due to variations in the data used for model training, choice of deep learning model, and parameterization process. We have performed rigorous and statistically significant empirical evaluation using average scores from human experts having diverse backgrounds and internal clustering stability measures. It presents as a baseline, and a promising start for future search results aggregation approaches.

Each research utilizes different techniques and mechanisms to provide information exploration and discovery. We have extracted the major parameters and their possible values for an in-depth comparison of our approach with existing state-of-the-art. To ease comprehension of these parameters, we have further categorized parameters according to their purpose, as displayed in the table 7. The provided functionalities are marked with the "+" symbol, whereas the missing functionalities are left blank.

Only 3 aspects are outlined below in the paragraph

Table 7 emphasizes the four significant aspects of discovery techniques. The first aspect is searching for search results including search type, search results granularity, and searching activity. The second aspect concerns data management, including information sources, instantiation of data-modal, and assembling mechanism. Finally, the third aspect is concerned with technical information retrieval aspects of the discovery and exploratory approaches, including media sources and information retrieval modal.

The most crucial factor in information discovery is flexibility in representing the information to avoid information overload. Most of the existing research solely relies on filtering capabilities but lacks in providing appropriate granularity control of the search results (18). Our approach provides three-level granularity; snippets, multimedia documents, and multimedia documents clusters. The data-modals employed by the existing researches are mostly centered around a specific domain and specific data. They mainly include the scientific domain having millions of literature as a dataset. Approaches providing real datasets were also primarily concerned with integrating a few verticals such as web and image (25). To enable our approach to be generic and applicable to all the domains and datasets, we presently use only real datasets to observe our approach's integrity even in the most variate and uncertain data coming from the search engines in real-time.

Information exploration and discovery is a long, non-trivial, and non-linear journey. To foster non-linear navigation of the search results, existing literature mostly instantiated a graph data-modal using either existing domain knowledge, such as ontologies (28; 37; 24), or using some generic similarity measures (42). Our approach uses domain-independent semantics and similarity measures to construct a non-linear graph to provide non-linear means of search results exploration and discovery.

Information management is also an essential factor in enabling information discovery and a compelling exploration of the search results. Numerous information management approaches organize and present the users' search results, increasing their cognitive abilities. These approaches include linear and non-linear browsing of information and summarization. However, previously, these were implemented as disjoint components, combined on a single interface (20). Our approach unifies all of the specific techniques and encapsulates it in a single component. With our generic information discovery architecture based on a strong theoretical background and promising empirical evaluation results, we hope to provide a new baseline for future researches on relational aggregated search and search engines alike.

CONCLUSION & FUTURE WORK

In this research, we proposed a generic discovery architecture using multimedia search engine results. A brief discussion on information exploration and theoretical discovery background was provided, and an architectural solution was formalized and instantiated. Before this work, the exploration and discovery of information on the web search engine were leveraged using traditional heuristics. We identified potential gaps and issues in the current general web search engine approach. To overcome these issues, we presented a new baseline using search results aggregation. Our approach was employed using state-of-the-art sentence embeddings. We bridged the gap between the abundant multimedia contents by encapsulating semantically multimedia artifacts in multimedia documents and summarizing them. Moreover, we eased the navigation problem in the search results space by grouping multimedia documents in semantically similar patches.

The proposed discovery architecture emphasizes all the aspects of the discovery, including information exploration and lookup. We supported information exploration by providing the nonlinear proximal navigation and exploration support through the instantiation of a complex graph and lookup searches through a semantically fully-blended ordered linear search results list. Finally, a comprehensive empirical evaluation was presented. The empirical evaluation out-performed previous aggregation approaches at all granularity levels of aggregation provided in this research. To the best of our knowledge, our approach is the first to be assessed comprehensively from the system and the user perspective on the dataset and the queries obtained from the user and the search engine, respectively, in real-time.

In the future, we look forward to providing a comprehensive usability perspective of architecture involving an even broader audience with extremely varied backgrounds and experiences with more focus on human aspects, including user interfaces. We have intentions to provide the adaptable clustering of multimedia documents by considering the users' diverse information need and information-seeking behavior. We are interested in exploiting various nonlinear data models in the enhanced discovery of aggregated multimedia based document search results in real-time scenarios.

REFERENCES

[1] Achsas, S. et al. (2019). An analysis study of vertical selection task in aggregated search. *Procedia computer science*, 148:171–180.

[2] Achsas, S. and Nfaoui, E. H. (2018). Improving relational aggregated search from big data sources using stacked autoencoders. *Cognitive Systems Research*, 51:61–71.

[3] Ali, S. and Gul, S. (2016). Search engine effectiveness using query classification: a study. *Online Information Review*.

[4] Athukorala, K., Głowacka, D., Jacucci, G., Oulasvirta, A., and Vreeken, J. (2016). Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, 67(11):2635–2651.

[5] Bakrola, D. and Gandhi, S. (2016). Enhancing web search results using aggregated search. In *Advances in Intelligent Systems and Computing*, volume 409, page 675–688. Springer Verlag.

[6] Baldonado, M. Q. W. and Winograd, T. (1996). Sensemaker: an information-exploration interface supporting the contextual evolution of a user's interests. In *CHI'97*.

[7] Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424.

[8] Batrinca, B. and Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, 30(1):89–116.
[9] Belkin, N. J., Oddy, R. N., and Brooks, H. M. (1982). Ask for information retrieval: Part i. background and theory. *Journal of documentation*, 38(2):61–71.
[10] Bianchi-Berthouze, N., Katsumi, N., Yoneyama, H., Bhalla, S., and Izumita, T. (2003). Supporting the interaction between user and web-based multimedia information. In *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, pages 593–596. IEEE.
[11] Bron, M., Van Gorp, J., Nack, F., Baltussen, L. B., and de Rijke, M. (2013). Aggregated search interface preferences in multi-session search tasks. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, page 123–132. ACM.
[12] Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.
[13] Campbell, P. (2013). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior (3rd ed.)*, volume 34. Emerald Group Publishing.
[14] Craswell, N., Campos, D., Mitra, B., Yilmaz, E., and Billerbeck, B. (2020). Orcas: 18 million clicked query-document pairs for analyzing search. *arXiv preprint arXiv:2006.05324*.
[Davies] Davies, D. et dw bouldin (1979). a cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell*, 1(2).
[16] Degbelo, A. and Teka, B. B. (2019). Spatial search strategies for open government data: A systematic comparison. In *Proceedings of the 13th Workshop on Geographic Information Retrieval*, pages 1–10.
[17] Deldjoo, Y., Schedl, M., Cremonesi, P., and Pasi, G. (2018). Content-based multimedia recommendation systems: definition and application domains.
[18] di Sciascio, C., Sabol, V., and Veas, E. E. (2016). Rank as you go: User-driven exploration of search results. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, page 118–129.
[19] Elzein, N. M., Majid, M. A., Hashem, I. A. T., Yaqoob, I., Alaba, F. A., and Imran, M. (2018). Managing big rdf data in clouds: Challenges, opportunities, and solutions. *Sustainable Cities and Society*, 39:375–386.
[20] Fung, C. C. and Thanadechteemapat, W. (2010). Discover information and knowledge from websites using an integrated summarization and visualization framework. In *2010 Third International Conference on Knowledge Discovery and Data Mining*, page 232–235.
[21] Gäde, M., Hall, M., Huurdeman, H., Kamps, J., Koolen, M., Skov, M., Toms, E., and Walsh, D. (2015). Supporting complex search tasks. In *European Conference on Information Retrieval*, pages 841–844. Springer.
[22] Google (2020). tidal waves - google search.
[23] Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for information Science*, 43(9):602–615.
[24] Kanjanakuha, N., Janecek, P., and Techawut, C. (2019). The comprehensibility assessment of visualization of semantic data representation (vsdr) reflecting user capability of knowledge exploration and discovery. In *Proceedings of the 2019 7th International Conference on Computer and Communications Management*, page 195–199.
[25] Kerne, A., Koh, E., Dworaczyk, B., Mistrot, J. M., Choi, H., Smith, S. M., Graeber, R., Caruso, D., Webb, A., Hill, R., and et al. (2006). Combinformation: A mixed-initiative system for representing collections as compositions of image and text surrogates. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, volume 2006, page 11–20.
[26] Kerne, A. and Smith, S. M. (2004). The information discovery framework. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, page 357–360.
[27] Kerne, A., Smith, S. M., Koh, E., Choi, H., and Graeber, R. (2008). An experimental method for measuring the emergence of new ideas in information discovery. *International Journal of Human-Computer Interaction*, 24(5):460–477.
[28] Khalili, A., Van Andel, P., Van Den Besselaar, P., and De Graaf, K. A. (2017). Fostering serendipitous knowledge discovery using an adaptive multigraph-based faceted browser. In *Proceedings of the Knowledge Capture Conference, K-CAP 2017*, page 15.
[29] Klouche, K., Ruotsalo, T., Cabral, D., Andolina, S., Bellucci, A., and Jacucci, G. (2015). Designing

for exploratory search on touch devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 4189–4198, New York, NY, USA. Association for Computing Machinery.

[30] Kopliku, A. (2009). Aggregated search: From information nuggets to aggregated documents. In *CORIA*, pages 495–502.

[31] Kopliku, A., Damak, F., Pinel-Sauvagnat, K., and Boughanem, M. (2011). Interest and evaluation of aggregated search. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*, page 154–161. IEEE Computer Society.

[32] Kopliku, A., Pinel-Sauvagnat, K., and Boughanem, M. (2014). Aggregated search: A new information retrieval paradigm. *ACM Computing Surveys*, 46(3).

[33] Krishnamurthy, Y., Pham, K., Santos, A., and Freire, J. (2016). Interactive web content exploration for domain discovery. *pdfs.semanticscholar.org*.

[34] Kumar, V. and Ogunmola, G. A. (2020). Web analytics for knowledge creation: a systematic review of tools, techniques, and practices. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, 10(1):1–14.

[35] Lewandowski, D. (2008). Search engine user behaviour: How can users be guided to quality content? *Information Services and Use*, 28(3-4):261–268.

[36] Li, X., Liu, Y., Cai, R., and Ma, S. (2017). Investigation of user search behavior while facing heterogeneous search services. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 161–170.

[37] Lisena, P., Troncy, R., Todorov, K., and Achichi, M. (2017). Modeling the complexity of music metadata in semantic graphs for exploration and discovery. In *Proceedings of the 4th International Workshop on Digital Libraries for Musicology*, page 17–24. ACM.

[38] Marchionini, G. (1997). *Information seeking in electronic environments*. Number 9. Cambridge university press.

[39] Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46.

[40] Oussous, A., Benjelloun, F. Z., Ait Lahcen, A., and Belfkih, S. (2018). Big Data technologies: A survey.

[41] Pirolli, P. and Card, S. (1999). Information foraging. *Psychological review*, 106(4):643.

[42] Rashid, U. and Bhatti, M. A. (2017). A framework to explore results in multiple media information aggregated search. *Multimedia Tools and Applications*, 76(24):25787–25826.

[43] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

[44] Ruotsalo, T., Jacucci, G., Myllymäki, P., and Kaski, S. (2015a). Interactive intent modeling: information discovery beyond search. *Commun. ACM*, 58(1):86–92.

[45] Ruotsalo, T., Peltonen, J., Eugster, M. J., Głowacka, D., Floréen, P., Myllymäki, P., Jacucci, G., and Kaski, S. (2018a). Interactive intent modeling for exploratory search. *ACM Transactions on Information Systems (TOIS)*, 36(4):1–46.

[46] Ruotsalo, T., Peltonen, J., Eugster, M. J. A., Głowacka, D., Floréen, P., Myllymäki, P., Jacucci, G., and Kaski, S. (2018b). Interactive intent modeling for exploratory search. *ACM Trans. Inf. Syst.*, 36(4).

[47] Ruotsalo, T., Peltonen, J., Eugster, M. J. A., Głowacka, D., Reijonen, A., Jacucci, G., Myllymäki, P., and Kaski, S. (2015b). Scinet: Interactive intent modeling for information discovery. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 1043–1044.

[48] Russell, D. M., Stefik, M. J., Pirolli, P., and Card, S. K. (1993). The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, page 269–276.

[49] Russell-Rose, T. and Tate, T. (2012). *Designing the search experience: The information architecture of discovery*. Newnes.

[50] Savolainen, R. (2018). Berrypicking and information foraging: Comparison of two theoretical frameworks for studying exploratory search. *Journal of Information Science*, 44(5):580–593.

[51] Sullivan, D. (2020). Meet the new google look its colorful, useful "search options" column.

[52] Sushmita, S., Joho, H., and Lalmas, M. (2009). A task-based evaluation of an aggregated search interface. In *International Symposium on String Processing and Information Retrieval*, page 322–333.

Springer.
[53] Sushmita, S., Joho, H., Lalmas, M., and Villa, R. (2010). Factors affecting click-through behavior in aggregated search interfaces. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, page 519–528. ACM.
[54] Sushmita, S., Lalmas, M., and Tombros, A. (2008). Using digest pages to increase user result space: Preliminary designs. In *SIGIR 2008 Workshop on Aggregated Search, Singapore*. Citeseer.
[55] Tablan, V., Bontcheva, K., Roberts, I., and Cunningham, H. (2015). M{\'\i}mir: An open-source semantic search framework for interactive information seeking and discovery. *Journal of Web Semantics*, 30:52–68.
[56] Taheri, S., Vedienbaum, A., Nicolau, A., Hu, N., and Haghighat, M. R. (2018). Opencv. js: Computer vision processing for the open web platform. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pages 478–483.
[57] Taramigkou, M., Apostolou, D., and Mentzas, G. (2017). Supporting creativity through the interactive exploratory search paradigm. *International Journal of Human–Computer Interaction*, 33(2):94–114.
[58] Tseng, L. C. J., Tjondronegoro, D., and Spink, A. (2009). Analyzing web multimedia query reformulation behavior. *ADCS 2009 - Proceedings of the Fourteenth Australasian Document Computing Symposium*, page 118–125.
[59] Wani, M. A. and Riyaz, R. (2016). A new cluster validity index using maximum cluster spread based compactness measure. *International Journal of Intelligent Computing and Cybernetics*.
[60] White, R. W., Kules, B., Drucker, S. M., et al. (2006). Supporting exploratory search, introduction, special issue, communications of the acm. *Communications of the ACM*, 49(4):36–39.
[61] White, R. W. and Roth, R. A. (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, 1(1):1–98.
[62] Zhang, W., Deakin, J., Higham, N. J., and Wang, S. (2018). Etymo: A new discovery engine for ai research. In *Companion Proceedings of the The Web Conference 2018*, page 227–230.