Volume 5, Issue 4, April 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Integrating News Verticals into Web Search Results

[1]Mohammad Ubaidullah Bokhari, [2]Mohd. Kashif Adhami

[1]Chairman, Dept. of Computer Sc. Aligarh Muslim University, Uttar Pradesh, India

[2]Research scholar, Dept. of Computer Sc. Aligarh Muslim University, Uttar Pradesh, India

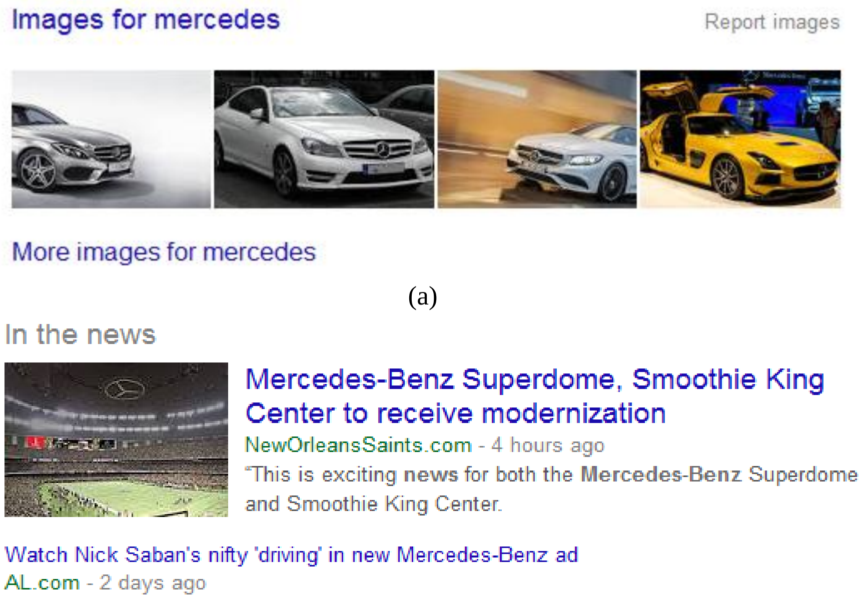
Abstract*— Aggregated search refers to the integration of content from different information sources, so-called verticals (image, video, blog, news etc.), into web search results. Aggregation improves search when the user has specific intent but may not be aware of or desire vertical search. 'Verticals' are the specialized corpora containing content for any specific domain. In this paper we address the job of integrating the search results from a news vertical into web search results and discuss how this integration contributes to the effectiveness of news search. Due to dynamic nature of news, integrating news verticals to the core web results is particularly challenging, especially decision, regarding 'when' and 'where' to add news vertical , changes with time. All major search engines are now doing some levels of aggregated search. So we also present an overview of current developments in aggregated search.

***Keywords*— News verticals, News search engines, Blended search, Resource selection and Universal search.**

I. Introduction

The main goal of web search engines is to display links to relevant web pages for the query submitted by the users. Recently, search engines have extended their services to include search on specialized collections of documents which focussed specific domains such as news, shopping or media. These specialized collections are called 'verticals'. Figure 1 shows some example verticals. There are two ways a user might submit his query to the web search engine, first a user may submit his query directly to the portal search engine (in case user might be unaware of a relevant vertical, or not willing to search a specific vertical) and secondly user may submit his query directly to a vertical search engine (in case when user believes that the relevant content exists in the vertical). In the case when a user have intent for news and he does not go for a vertical search engine or a news search engine and directly submits the query to the portal search engine then aggregated search result will satisfy his need.

There is a difference between standard web search and aggregated search. In core web searching, documents of the same nature are compared such as web pages or images and ranked according to their estimated relevance to the query. In case of aggregated search, documents of a different nature are compared, such as web pages against images and their relevance is estimated with respect to each other. Different algorithms are required to rank these heterogeneous information items. Also relevance of verticals differ with queries. For example, for the query 'black mercedes' it might make more sense to return, in addition to core web result, documents from one vertical such as image vertical than from another like news vertical. Thus the main challenge in aggregated search is how to identify and integrate relevant heterogeneous results for each given query into a single result page.



Mercedes-Benz Superdome, Smoothie King Center to receive modernization
NewOrleansSaints.com - 4 hours ago
"This is exciting news for both the Mercedes-Benz Superdome and Smoothie King Center.

Watch Nick Saban's nifty 'driving' in new Mercedes-Benz ad
AL.com - 2 days ago

(a)

(b)

Fig. 1 (a) Image vertical and (b) News vertical

Aggregated search has three main components: (1) *vertical representation,* concerned with how to represent verticals so that documents contained within and their type are identifiable, (2) *vertical selection*, concerned with how to select the relevant vertical from which relevant information can be retrieved, and (3) *result presentation,* concerned with how to assemble aggregated results so as to best layout the relevant information to the user.

II. Background and motivation

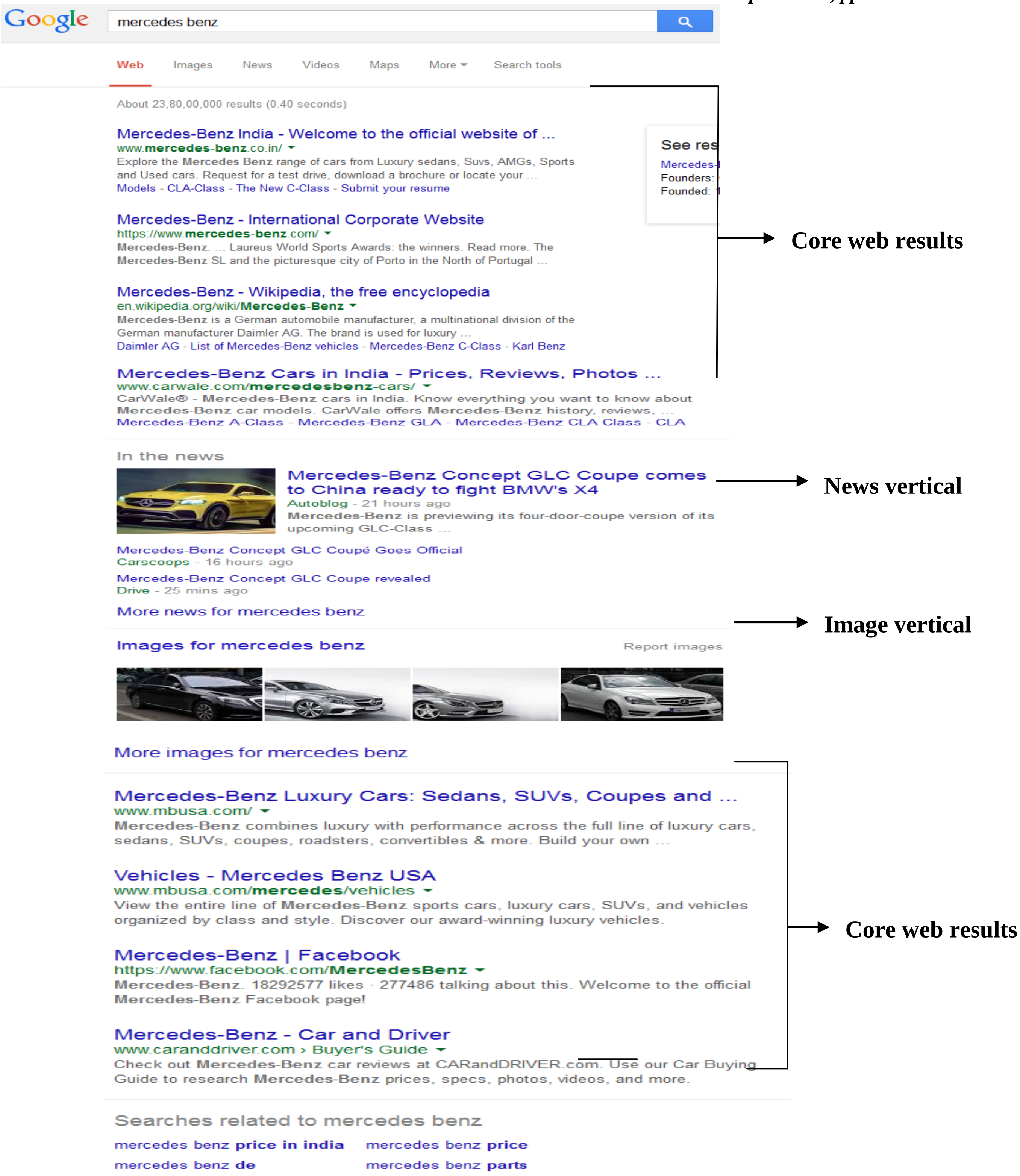
Aggregated search seeks diverse information across heterogeneous verticals. This diverse information searching is not new. Federated search ([1],[12],[13] and [14]) and metasearch [15-19] are techniques that aim to search and provide results from diverse sources. In federated search, a user after submitting the query, may select a number of sources, reffered to as resources, to search. These resources are often standalone systems like internet sources, corporate intranets, library catalogs etc. In case when not explicitly stated by the user, the federated search system has to identify the most relevant resources (resource selection) for searching. The given query is then sent to the selected resources. These resources then return results for that query to the federated search system. The system then decides which and how many results to retain. Finally these selected results are presented to the users. The returned results are often merged within one single ranked list, which can, if needed, can be separated. In case, when system return duplicate results, then it should be removed. Some of the examples of federated search are : Westlaw, FedStats and Funnelback.

An extensive survey of the federated search was given by Shokouhi and Si [1]. When federated search was brought to the web then two paradigms originated, metasearch and aggregated search. A metasearch engine is a search engine that sends the query to several different search engines and combines results from them and presents, to the user, a single merged ranked list. The metasearch engine operates on the fact that a more comprehensive search results can be obtained by combining results from different search engines as the web is too large for any one search engine to index. According to M. Lalmas in [2] metasearch engines were more popular 10-15 years ago and now the partial coverage of the web space is not an issue with current major search engines like Google, Yahoo and Bing. Also current search engines do not usually provide unlimited access to their search results to any third party application like metasearch engine probably because of incurred traffic loads and business models. Some of the examples of existing metasearch engines include Dogpile, Metacrawler and Search.Com.

The second paradigm, aggregated search, also provides information from different sources. These information sources are powered by dedicated vertical search engines which are mostly within the remit of the general web search engine instead of several and independent search engines, as in the case of metasearch engines. In aggregated search, the individual information sources retrieve from very different collections of documents like images, news and videos etc. A typical example of an aggregated search is shown in figure 2. It shows the result page for the query “mercedes bens” in google. The result page shows core web results blended with vertical results. Google introduced the idea of universal search first in 2007. The goal behind the use of aggregated search is to remedy the fact that most web users do not prefer to use vertical search and mostly go for portal search engines. According to the survey conducted by JupiterResearch (iProspect 2008)[20] in 2007-2008, 35% of the users do not use vertical search. But this does not mean that vertical search intent is not present within web search queries. The fact that queries can be answered from various verticals was shown by Arguello et al [3], who looked at 25,195 unique queries obtained from a commercial search engine query log. For analysing user vertical intent, human editors were instructed to assign between zero and six relevant verticals per query based on their best guess. About 26% of queries were assigned no relevant vertical and 44% were assigned a single relevant vertical and rest were assigned multiple relevant verticals because they were ambiguous in terms of vertical intent. Other works analysing vertical search intent were [4] and [5]. Table1 shows the summary of vertical search intent analysis works.

Table 1:Different works analyzing vertical search intent within web search queries

Work	**Total no. of queries**	**Method**	**Observations**
Arguello et al [3]	25,195	Human annotation, 0 to 6 relevant verticals per query	26%- no relevant verticals 44%- single relevant vertical 30%- multiple relevant verticals
Liu et al [4]	2,153	Analysed query logs from a commercial search engine.	12.3%-image search intent. 8.5%-video search intent. 13.1%-news search intention. 0.9%-blog search intention. 7.2%-book search intention.
Sushmitha et al [5]	13,279,076 log entries from Microsoft 2006 RFP dataset.	Analysed query log click-through in terms of vertical intent.	User do not often mix intents (if they do so, mostly used two intents). Users do not switch back and forth within intents. For video, news and map intents, different queries were submitted. For blog and Wikipedia intent, same query was used.



About 23,80,00,000 results (0.40 seconds)

Mercedes-Benz India - Welcome to the official website of ...
www.mercedes-benz.co.in/
Explore the Mercedes Benz range of cars from Luxury sedans, Suvs, AMGs, Sports and Used cars. Request for a test drive, download a brochure or locate your ...
Models - CLA-Class - The New C-Class - Submit your resume

Mercedes-Benz - International Corporate Website
https://www.mercedes-benz.com/
Mercedes-Benz. ... Laureus World Sports Awards: the winners. Read more. The Mercedes-Benz SL and the picturesque city of Porto in the North of Portugal ...

Mercedes-Benz - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Mercedes-Benz
Mercedes-Benz is a German automobile manufacturer, a multinational division of the German manufacturer Daimler AG. The brand is used for luxury ...
Daimler AG - List of Mercedes-Benz vehicles - Mercedes-Benz C-Class - Karl Benz

Mercedes-Benz Cars in India - Prices, Reviews, Photos ...
www.carwale.com/mercedesbenz-cars/
CarWale® - Mercedes-Benz cars in India. Know everything you want to know about Mercedes-Benz car models. CarWale offers Mercedes-Benz history, reviews, ...
Mercedes-Benz A-Class - Mercedes-Benz GLA - Mercedes-Benz CLA Class - CLA

In the news

Mercedes-Benz Concept GLC Coupe comes to China ready to fight BMW's X4
Autoblog - 21 hours ago
Mercedes-Benz is previewing its four-door-coupe version of its upcoming GLC-Class ...

Mercedes-Benz Concept GLC Coupé Goes Official
Carscoops - 16 hours ago

Mercedes-Benz Concept GLC Coupe revealed
Drive - 25 mins ago

More news for mercedes benz

Images for mercedes benz

More images for mercedes benz

Mercedes-Benz Luxury Cars: Sedans, SUVs, Coupes and ...
www.mbusa.com/
Mercedes-Benz combines luxury with performance across the full line of luxury cars, sedans, SUVs, coupes, roadsters, convertibles & more. Build your own ...

Vehicles - Mercedes Benz USA
www.mbusa.com/mercedes/vehicles
View the entire line of Mercedes-Benz sports cars, luxury cars, SUVs, and vehicles organized by class and style. Discover our award-winning luxury vehicles.

Mercedes-Benz | Facebook
https://www.facebook.com/MercedesBenz
Mercedes-Benz. 18292577 likes · 277486 talking about this. Welcome to the official Mercedes-Benz Facebook page!

Mercedes-Benz - Car and Driver
www.caranddriver.com › Buyer's Guide
Check out Mercedes-Benz car reviews at CARandDRIVER.com. Use our Car Buying Guide to research Mercedes-Benz prices, specs, photos, videos, and more.

Searches related to mercedes benz

mercedes benz price in india
mercedes benz price
mercedes benz de
mercedes benz parts

Fig. 2 Aggregated search result for 'mercedes benz' query in Google.

The JupiterResearch (iprospect 2008) [20] survey also showed that within one year of major search engines providing users with aggregated search results, a larger number of users clicked on vertical search result types within the general search results, against when the verticals were searched directly. According to them, news results were the most clicked vertical results within aggregated search, and even more than using directly the vertical news search, users click them more than twice as much within aggregated search. Similar trend can be obsereved in CompScore report [6]. Thus we can say that despite users limited use of vertical search engines, it is important for search engines to integrate relevant verticals in their result pages.

III. AGGREGATED SEARCH TASKS

Aggregated search consists of three main tasks: vertical representation, vertical selection and result presentation.

A. Vertical Representation

To select and aggregate relevant vertical(s) for each submitted query, an aggregated search engine needs to know about the content of verticals to ensure that submitted query is passed to the appropriate vertical. For this, the aggregated search system keeps a representation of each of its verticals. Vertical representation task is similar to the resource representation task in federated search [1]. A vertical representation can be built using techniques from federated search working with cooperative resources. Several techniques have been proposed in [1]. A technique reported in [1] is the generation of vertical representations from a subset of documents, so-called sampled documents.

Two other sampling approaches were reported in Arguello et al [3]. In the first method, documents are sampled directly from the vertical and the second method used the external sources for sampling.

Query-based sampling was used to directly sample from the vertical. This can be explained in the context of federated search. First, an initial query is used to retrieve documents from the resource, which are then used to built the initial representation of the resource. Then new sampling query is selected from the representation. The generated resource representation and sampling queries are derived from retrieved documents. In [1], it was shown that we get better performance when high-frequency queries were used for sampling the documents, than when derived from sampled documents themselves. Similarly, in the context of aggregated search, high frequency queries are used for sampling documents [3]. Queries issued directly to the vertical represent explicit vertical intent and serves as a good candidate for sampling.

The alternative method is to sample from external sources, provided that the documents can be mapped directly to verticals. In [3] Arguello et al sampled from Wikipedia, making use of Wikipedia categories, either one or several, assigned to documents to map them to verticals. Sampling from Wikipedia can be helpful in providing better representation of text-impoverished verticals such as image and video verticals. In addition they have consistent format that are semantically coherent and on topic, so better coverage of the vertical content and more uniform representation across verticals can be achieved.

B. Vertical Selection

Vertical selection is the task of selecting the relevant verticals, either none, one or many) in response to a user query. Vertical selection make use of the vertical representations for selecting relevant vertical(s). This task can be treated similar to the resource selection task in federated search, where resource representation was used for selecting and merging various relevant resources for achieving diversity in information needs. Some of the prior approaches such as CVV[21], CORI[22], UUM[23] and KL divergence [24] consider sampled documents or collections as "large documents". These sampled documents or collections were given scores using document scoring techniques. These techniques made no distinction between documents and do not model the number of relevant documents in a collection. Verticals actually focus on specific types of documents (in terms of domain, media or genre [5]). One of the source of evidence for vertical selection can be query string, as users searching for a particular type of document, like "images of Toyota cars", may issue domain (or genre/media) specific words in the query. Secondly, the vertical search engines are being used explicitly by users, so query logs can be exploited for vertical selection. These two sources along with vertical representations have been reported in [3], [7] and [8]. Machine learning algorithms/techniques were used to built models of verticals based on features extracted from vertical representations. More precisely we can say Arguello et al [3] used features from vertical representations such as vertical ranking scores similar to those produced by ReDDE[13] in federated search [1]. Query string features were based on rule triggers by word occurrences in the query. These rules mapped words to appropriate verticals, such as "car" to the autos vertical. Features for the vertical query logs correspond to the query likelihood generated from a unigram language model built from the vertical query logs. Findings showed that ranking verticals by the query likelihood was the best single evidence to select a vertical. Using rules mapping query strings to verticals also led to the significant improvement in vertical selection, particularly in the situations where no training data is available for a vertical. In [8] the technique of machine adaptation was employed to learn how models learned for a vertical with training data could be ported to other verticals having no training data. Li et al [10] classified queries into two classes of vertical intent namely product and job. The vertical was selected if it satisfies user's job or product intent. Diaz [11] predicted the integration of news vertical always above the web result, i. e., at the top position. He used correctly predicted clicks and skips as evaluation measure for predicting the preference.

C. Result Presentation

There are two main types of result page design in aggregated search. The first one is known as *blended* design where results from different verticals are blended into a single list. In the second one, results from each vertical are presented in a separate panel and referred to as *non-blended* design. The blended design is the most common way to present results in aggregated search and is applied by Google universal search and many other search engines. With the blended design, the relevance is the main ranking criteria within and across verticals. Including web search result, the results from the same vertical are slotted together and the entire slot is ranked with respect to other slots.

The blended aggregated search have been investigated in [4] and [9]. Liu et al [4] used machine learning techniques to estimate probabilities such as a document in a vertical being relevant, a vertical being relevant, etc., for probabilistic model. All the documents were than ranked, from the standard web and vertical results, using the resulting probabilistic document scores, into a one single list. An increase in performance was observed but the quality of the blended search results was not really evaluated.

Another work from Ponnuswami et al [9] investigated blended aggregated search. Their work focussed on placement of selected relevant verticals at the right position in the blended interface. Machine learning techniques were used on training data based on elicited pairwise preferences between groups of results (group of standard results and a group of vertical results). While ranking, group of web results were compared with group of vertical results. A group of vertical results is placed at a slot if the score is higher than some threshold employed to guarantee specific coverage for verticals at each slot. This was the first work publishing detailed account on composition of blended result page.

In non-blended integration results from each vertical is presented in a separate panel. This panel is also referred to as 'tile'. Some of the examples using such designs are alpha yahoo, kosmix and Naver. The main web search results are displayed on the left side and within the largest panel. Results from different panels, does not have any relationship with each other and also the placement of the various panels are predefined.

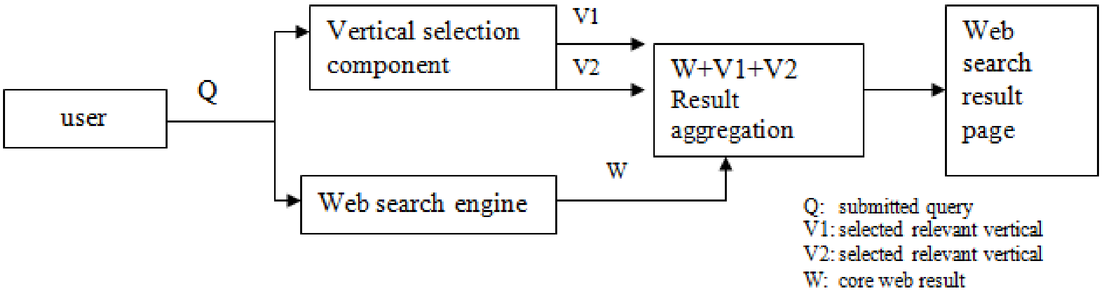

user
Q
Vertical selection component
V1
V2
Web search engine
W
W+V1+V2 Result aggregation
Web search result page
Q: submitted query
V1: selected relevant vertical
V2: selected relevant vertical
W: core web result

Fig. 3 A high level overview of results aggregation.

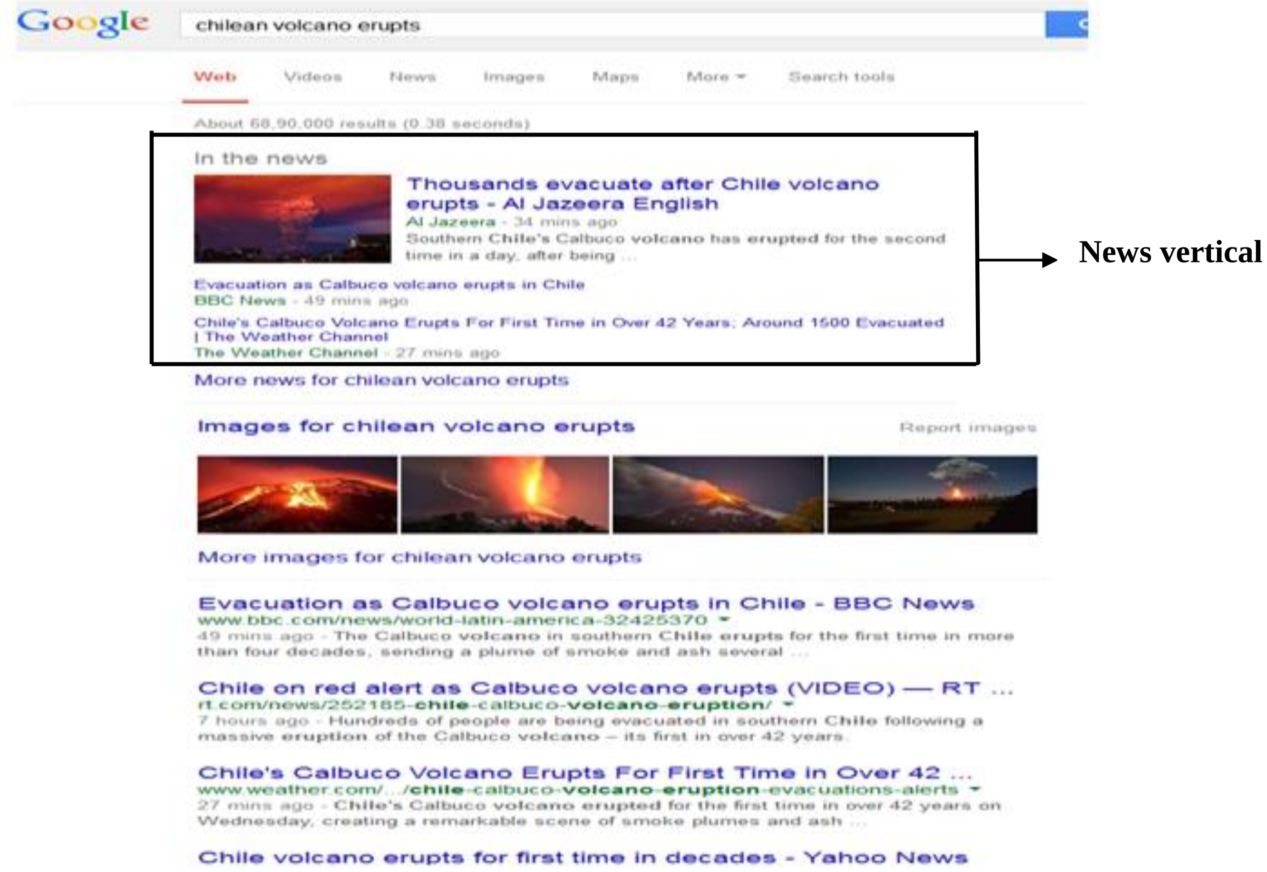
IV. Integration of News verticals

This section discusses the integration of news verticals into the web search results and its impact on effectiveness of news searching. The users having news intent submits their query usually in either of the two ways, first the users can directly use the vertical/news search engine and secondly they can enter their query into the portal web search engine interface. The latter case uttered the need for studying various factors associated with the selection of appropriate news vertical and its integration at the appropriate place, in other words *which* news vertical to integrate and *where* to integrate in the result page. Also we will review some eye tracking studies and relate it with the effective news searching using aggregated search. A web searcher may express explicit intent for vertical intent(e.g. delhi election news) or the searcher's intent may be implicit (e.g. delhi election). News results are presented above the top web result in a small box referred to as the news display or view. We experimented this with seven implicit queries in google search engine. Table 2 shows that news verticals are displayed, in most cases, above the web results. Figure 4 shows an example news display for the query 'chilean volcano erupts'. Due to dynamic nature of the news, given a query, the appropriate decision to integrate news content or not changes with time. These changes occur at two places. First in the news index, topics emerge and decay with respect to content production and secondly in query logs, news intent grows and decays against the content demand. A system which only models evolving topics in the news index may waste their modeling effort on topics never requested by the searchers or which are not newsworthy. Similarly a system which only models evolving query volume might not be able to separate queries requiring news display from those which are merely popular.

Table 2: Collection of implicit queries for determining the position of news vertical in result page.

S. No.	Submitted query	Date of incident	Position of the news vertical
1	Germanwings plane crash	31 march 2015	just next to the top result.
2	European champions league semifinals	22 april 2015	at the top.
3	Chilean volcano erupts	23 april 2015	at the top.
4	PM Modi's visit to Canada	17 april 2015	just next to the top result.
5	Tendulkar's last match	14 Nov 2013	No news vertical
6	Delhi polls 2015	14 feb 2015	No news vertical
7	Japan manglev train's world record	22 april 2015	at the top.

Diaz [11] presented a system which integrated both massive document and query approaches to modeling events. They trained a classifier to distinguish between newsworthy and non-newsworthy queries. For training a classifier for any task, one requires a training set. For classifying queries as whether deserving a news display human annotation is required but for humans making this decision requires knowledge about topical events being queried and also topical events being written at that time when query was issued and this task is extremely expensive and potentially unreliable. Diaz [11], to address this issue, defined several click-based metrics which allowed a system to be monitored and tuned without human annotation. They demonstrated that the feedback is sufficiently related to manual labels so as to allow it to be used as a surrogate for training.



chilean volcano erupts

About 68,90,000 results (0.38 seconds)

In the news

Thousands evacuate after Chile volcano erupts - Al Jazeera English
Al Jazeera - 34 mins ago
Southern Chile's Calbuco volcano has erupted for the second time in a day, after being ...

Evacuation as Calbuco volcano erupts in Chile
BBC News - 49 mins ago

Chile's Calbuco Volcano Erupts For First Time in Over 42 Years; Around 1500 Evacuated | The Weather Channel
The Weather Channel - 27 mins ago

More news for chilean volcano erupts

Images for chilean volcano erupts

More images for chilean volcano erupts

Evacuation as Calbuco volcano erupts in Chile - BBC News
www.bbc.com/news/world-latin-america-32425370
49 mins ago - The Calbuco volcano in southern Chile erupts for the first time in more than four decades, sending a plume of smoke and ash several ...

Chile on red alert as Calbuco volcano erupts (VIDEO) — RT ...
rt.com/news/252185-chile-calbuco-volcano-eruption/
7 hours ago - Hundreds of people are being evacuated in southern Chile following a massive eruption of the Calbuco volcano – its first in over 42 years.

Chile's Calbuco Volcano Erupts For First Time in Over 42 ...
www.weather.com/.../chile-calbuco-volcano-eruption-evacuations-alerts
27 mins ago - Chile's Calbuco volcano erupted for the first time in over 42 years on Wednesday, creating a remarkable scene of smoke plumes and ash ...

Chile volcano erupts for first time in decades - Yahoo News

Fig. 4 News vertical integrated at the top position for the query 'Chilean volcano erupts'.

V. Conclusions AND FUTURE WORK

Aggregated search refers to the integration of content from heterogeneous information sources called verticals into the web search results. Verticals are the specialized corpora containing content for any specific domain. In this paper we address the issue of integrating news verticals to the core web results and discussed how this integration contributes to the effectiveness of news search. For aggregating news search and web search, system should adapt to new events in response to document volume, query volume and click feedback. For this click feedback can be utilized to provide evidence to adaptively improve non-lexicographic baseline models, in other words determining newsworthiness relies on predicting the probability of a user clicks on the news display of a query. Secondly new methods for incorporating query similarity to improve performance on queries with high click through rate should be explored. More work can be done for gathering user feedback for low performing queries. We also presented an overview of current developments in aggregated search as now all major search engines are doing some level of aggregated search and we expect this to be rich and fertile area for future research.

References

[1] M. Shokouhi and L. Si. "Federated Search", Foundations and Trends in Information Retrieval, Vol 5(1), pp. 1-102, 2011.

[2] M. Melucci and R. Baeza-Yates. "Aggregated Search (Chapter-5 by M. Lalmas)", Advanced Topics in Information Retrieval, The Information Retrieval Series 33, Springer-Verlag, Berlin Hiedelberg, 2011.

[3] J. Arguello, F. Diaz, J. Callan and J. F. Crespo. "Sources of Evidence for Vertical Selection", In *Proc. of SIGIR'09,* Boston, Massachusetts, USA, 2009.

[4] N. Liu, J. Yan and Z. Chen. "A Probabilistic Model Based Approach for Blended Search", In proc. *of WWW,* Madrid, Spain, pp. 1075-1076, 2009.

[5] S. Sushmita, B. Piwowarski and M. Lalmas. "Dynamics of Genre and Domain Intents", In *Proc. of AIRS 2010,* Taipei, Taiwan, Information Retrieval Technology, pp. 399-409, 2010.

[6] E. Goodman and E. Feldblum. "Blended Search and the New Rules of Engagement", CompScore Report, 2010.

[7] F. Diaz and J. Arguello. "Adaptation of Offline Selection Predictions in Presence of User Feedback", In *Proc. of SIGIR 2009,* pp. 323-330, 2009.

[8] J. Arguello, F. Diaz and J. F. Paiement. "Vertical Selection in Presence of Unlabled Verticals", In *Proc. of SIGIR 2010,* pp. 691-698, 2010.

[9] A. K. Ponnuswami, K. Pattabiraman, Q. Wu, R. Gilad-Bachrach and T. Kanungo. "On Composition of a Federated Web Search Result Page: Using Online Users to Provide Pairwise Preference for Heterogeneous Verticals", In *Proc. of WSDM2011*, Hong Kong, 2011.

[10] X. Li, Y. Yang and A. Acero. "Learning Query Intent from Regularized Click Graphs", In *Proc. of SIGIR 2008,* pp. 339-346, 2008.

[11] F. Diaz. "Integration of News Content into Web Results", In *Proc. of WSDM 2009,* Barcelona, Spain, 2009.

[12] F. Diaz, M. Lalmas and M. Shokouhi. "From Federated to Aggregated Search", In *Proc. of SIGIR 2010,* pp. 910, 2010.

[13] L. Si and J. Callan. "Relevant Document Distribution Estimation Method for Resource Selection", In *Proc. of SIGIR 2003,* pp. 298-305, 2003.

[14] L. Si, R. Jin, J. Callan and P. Ogilvie. "A Language Modeling Framework for Resource Selection and Result Merging", In *Proc. of CIKM 2002,* pp. 391-397, 2002.

[15] L. Gravano, C. Chang, H. Garca-Molina and A. Paepcke. "STARTS: Stanford Proposai for Internet Metasearching", In *Proc. of SOGMOD 1997,* pp. 207-218, 1997.

[16] W. Y. Meng and C. Yu. "Advanced Metasearch Engine Technology", Morgan & Claypool Publishers, ISBN 1608451925, 2010.

[17] E. Selberg and O. Etzioni. "Multi-Service Search and Comparison using Metacrawler", In *Proc. of WWW 1995,* Boston, MA: Oreilly, 1995.

[18] E. Selberg and O. Etzioni. "The MetaCrawler Architecture for Resource Aggregation on the Web", IEEE Expert, vol: 12(1), pp. 8-14, 1997.

[19] S. Gauch and G. Wang. "Information Fusion with ProFusion", In *Proc. of the first web conference of the web society,* San Francisco, CA, USA, pp. 174-179, 1996.

[20] .http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm, iProspect Blended Search Results Study.

[21] B. Yuwono and D. L. Lee. "Server Ranking for Distributed Text Retrieval Systems on the Internet", In *Proc. of DASFAA 1997,* pp. 1-50, World Scientific Press, 1997.

[22] J. P. Callan, Z. Lu and W. B. Croft. "Searching Distributed Collections with Inference Networks", In *Proc. of SIGIR'95,* pp. 21-28, 1995.

[23] L. Si and J. Callan. "Unified Utility Maximization Framework for Resource Selection", In *Proc. of ACM CIKM'04,* Washington DC, USA, pp. 32-41, 2004.

[24] J. Xu and W. B. Croft. "Cluster-Based Language Models for Distributed Retrieval", In *Proc. of SIGIR'99,* pp. 254-261, 1999.