

# Technical Disclosure Commons

---

## Defensive Publications Series

---

May 2020

## Using Fresh Web Content To Generate Subscribable Search Notifications

Yew Jin Lim

James Kunz

Jon Dreyfus

Mauricio Monico

Tao Feng

*See next page for additional authors*

Follow this and additional works at: [https://www.tdcommons.org/dpubs\\_series](https://www.tdcommons.org/dpubs_series)

---

### Recommended Citation

Lim, Yew Jin; Kunz, James; Dreyfus, Jon; Monico, Mauricio; Feng, Tao; and Palem, Maya, "Using Fresh Web Content To Generate Subscribable Search Notifications", Technical Disclosure Commons, (May 29, 2020) [https://www.tdcommons.org/dpubs\\_series/3280](https://www.tdcommons.org/dpubs_series/3280)



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

---

**Inventor(s)**

Yew Jin Lim, James Kunz, Jon Dreyfus, Mauricio Monico, Tao Feng, and Maya Palem

## **Using Fresh Web Content To Generate Subscribable Search Notifications**

### **ABSTRACT**

Content subscriptions are a useful way for a search engine to maintain user engagement. However, generating and providing content updates regarding topics of interest based on knowledge repositories can be expensive. Further, updates generated in this manner can be too infrequent, providing an unsatisfactory user experience. This disclosure describes the automated analysis and use of fresh web content related to topics of interest to generate and deliver updates to subscribers. Fresh web content is analyzed, e.g., using a natural language classifier, to determine the topic and type of content. Updates are generated based on specific content types known to be of user interest and are provided to the user. The use of fresh content and filtering to certain content types ensures that the updates are timely, relevant, and interesting to the user.

### **KEYWORDS**

- Content subscription
- Search subscription
- Topic of interest
- Content type
- Content inventory
- Fresh content
- New content
- Search engine

## BACKGROUND

Users utilize search engines to obtain information about topics of their interest such as their favorite artists, sports teams, TV shows, etc. Some search engine providers provide users with an option to subscribe to ongoing notifications regarding their topics of interest. However, for subscriptions to be valuable to users, the cadence at which results are delivered to the users, is important - new and relevant results that are likely to be of interest to users need to be identified and delivered in a timely manner.

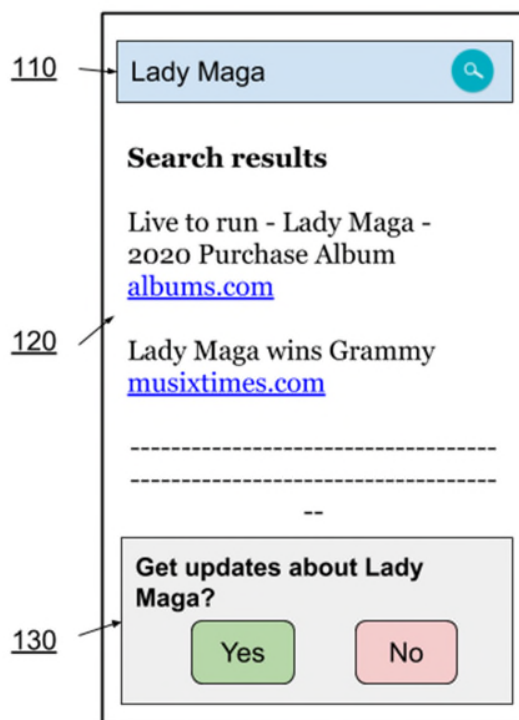
Knowledge repositories that include structured or semi-structured content (e.g., knowledge about an artist can include artist bio, information about new album releases, live events featuring the artist, etc.) are a useful source of information. However, such repositories are expensive to update and are updated relatively infrequently (e.g., when the artist releases a new album, or launches a new tour), and subscriptions based on such information are likely to be infrequent. Also, such repositories are hard to generalize and scale to different types of entities that are of interest to users.

## DESCRIPTION

For subscriptions to topics of interest to be useful, scalable and relevant sources of content that can engage users are necessary. This disclosure describes techniques to generate relevant updates based on fresh web content relating to the topics of interest indicated by users.

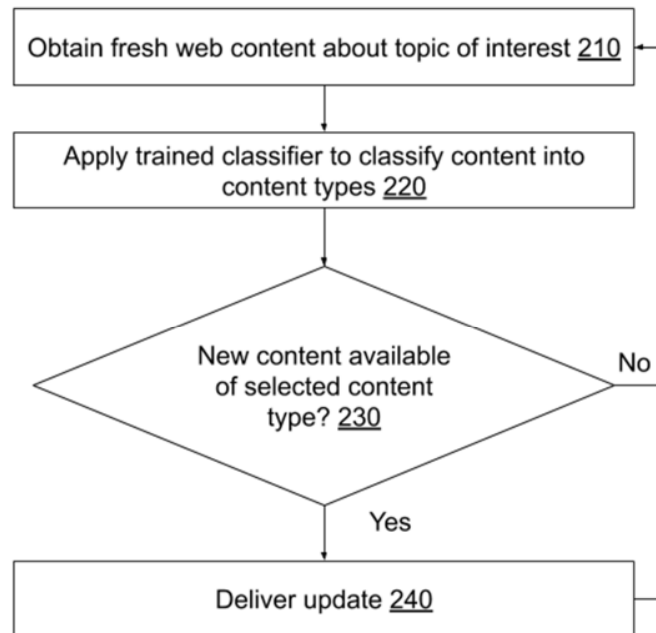
Such updates are selected based on the types of content users are likely to be interested in. For example, for a musical or video artist, the content types of interest may include interviews with the artist, awards or nominations that the artist receives, reviews of the artist's work, social media updates by the artist, etc. while content types that are not likely of interest to users may include gossip about personal lives of the artist. Fresh web content is analyzed using

a natural language classifier to classify the web content into the different content types. Updates on the topics are delivered when the type of content is deemed likely of interest.



**Fig. 1: Users subscribe to notifications based on their search results**

Fig. 1 illustrates an example of a user subscription to a topic of interest based on a user search. In this illustrative example, a user performs a search (110) using their computing device. In addition to the search results (120), the user is also provided with an option to subscribe (130) to obtain subsequent notifications/updates relating to the topic of interest.



**Fig. 2: A classifier is used to identify content of interest**

Fig. 2 illustrates an example workflow to identify content of interest based on web content relating to a topic, per techniques of this disclosure. Fresh web content relating to the topic of a user's interest is obtained (210), e.g., using a web crawler. Natural language processing (NLP) techniques are utilized to classify the obtained content into various content types. For example, a trained classifier is applied (220) to classify each obtained document into particular content types, e.g., "interview," "review," "award," "gossip," etc. The classifier can be trained using training data that includes web content with human assigned labels of content types or other training methods.

If new content is available that is of a selected content type (230), an update is generated and delivered to users (240) that have configured a subscription for the topic. For example, the update can be delivered as a notification. The content types that are of user interest can be determined, e.g., based on aggregated interaction data regarding past updates

**Example:** Fresh web content about the topic of interest, e.g., “Lady Gaga” is identified and processed using natural language classifiers to identify the type of content, e.g., interview with Lady Gaga (“Interview”), review of new album by Lady Gaga (“Review”), news about Lady Gaga’s shopping trip (“Gossip”) etc. When fresh web content of certain types, e.g., interview, review, is available, an update is delivered to user’s that subscribe to the topic “Lady Gaga.”

The techniques of this disclosure enable search subscriptions that provide updates at a sufficient cadence, while ensuring that the delivered update is timely and useful for the user. Compared with subscriptions that rely on structured/unstructured content repositories, the engineering cost of generating and delivering updates is lower, e.g., since automated natural language classifiers can produce the updates.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user’s topics of interest, subscriptions, a user’s preferences, a user’s location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user’s identity may be treated so that no personally identifiable information can be determined for the user, or a user’s geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

## CONCLUSION

This disclosure describes the automated analysis and use of fresh web content related to topics of interest to generate and deliver updates to subscribers. Fresh web content is analyzed, e.g., using a natural language classifier, to determine the topic and type of content. Updates are generated based on specific content types known to be of user interest and are provided to the user. The use of fresh content and filtering to certain content types ensures that the updates are timely, relevant, and interesting to the user.