

Survey- Similarity Project

Users who search for products online usually look for other products offered by the website and compare similar offers, hot dills, etc. To better support users in their product searching process, websites often provide recommendation mechanisms for suggesting similar and complementary products.

In this project we propose several product - offering methods and compare their effectiveness. In particular, the proposed methods are aimed at measuring the similarity between products and suggesting the most similar products to a searched product. We apply these methods in the context of tourism, where products are flights to different destinations.

We propose four methods:

1. **Based on users' search history** – similarity between two products is measured based on the number of times these products were searched for one after the other by users of the website. To implement this, we will build a weighted graph, where the nodes represent searched products (recorded on the website user-search logs) and the weights on edges represent the number of times the two connected products were searched for together.
2. **Based on product descriptions** – we will measure similarity between two products based on the similarity of their descriptions. To operationalize this, we will develop a crawler that will scrape data from the internet.
3. **Based on product images** – we will use deep learning methods to find the destination with the best offer/fit based on the similarity of photos from other destinations and compare them to the user searches.
4. **Combination** of the above methods – similarity between products will be calculated based on a combination of the above methods.

At the end of the project, we answer the following questions:

1. Can we use any of the methods mentioned above to construct an effective similarity measure between products.
2. Which method gives better results.
3. Check if using a combination of the methods will provide a better offer.

To compare the performance of each of the methods, we will use Amazon Mechanical Turks (AMT) which provides a platform for easy collecting answers or opinions from actual people. We will let people to fill out questionnaires about different destinations and similarities between them, these questionnaires will help us obtain the “ground truth” regarding the similarity between different destination. Then, the similarities measured by the different methods will be compared to this ground truth and the best method will be identified.

Techniques that we will use in the project:

* **Deep learning** – for finding the best offer for the user, we will use Tensorflow – Tensorflow is a platform that we will use to train a deep neural network using data collected using crawlers.

* **Crawler** – we will implement a web tool to retrieve information from the internet. We will develop two crawlers:

1. A crawler that collects images of destinations from Google Images.
2. A crawler that collects data from Lonely Planet, a website that provides detailed information on destinations for tourists.

* **Algorithms** – we will use different algorithms throughout the project including graph algorithms, deep learning algorithms, etc.

* **AMT** – we will use Amazon Mechanical Turks for collecting similarity information from people via questionnaires. This information will be used for evaluating the performance of each of the proposed methods for measuring the similarity between products.