Developing the Key Attributes for Product Matching Based on the Item's Image Tag Comparison

Olga Cherednichenko [0000-0002-9391-5220], Olha Yanholenko [0000-0001-7755-1255], Olga Kanishcheva [0000-0002-9035-1765]

National Technical University "Kharkiv Polytechnic Institute",
2, Kyrpychova str., 61002 Kharkiv, Ukraine
olha.cherednichenko@gmail.com, olga.yan26@gmail.com,
kanichshevaolga@gmail.com

Abstract. With the constant growth of the number of products on e-marketplaces, buyers feel hard to find and choose items that would satisfy all their needs and expectations. Search and filtering algorithms of recommender systems, although are striving to help users, still fail quite often due to incomplete and inaccurate description of items. The given work suggests to combine analysis of both item description and item image in order to construct groups of similar items. Since a person can define whether two items are similar or not looking at two images and a brief description, it is suggested to form a set of similar items based on users' judgments and then to extract the core of keywords for the specific type of products. Further, it is proposed to use the given core to evaluate the similarity of any new item added to the definite group. The case study deals with the building of the core of keywords for sneakers. The developed key attributes allow matching the items with a high precision, thus, proving the effectiveness of the method of the core construction.

Keywords: E-commerce; Item's Images, Similarity Items, Image Similarity, Images Matching, Tag Similarity, Key Attributes

1 Introduction

E-commerce marketplaces are the inexhaustible source of goods that may satisfy the needs of any exacting client. But often, the task to find the necessary product becomes a real challenge for a customer. While the sellers are allowed to create multiple item records for the same product, the customers see the endless lists of recommended items that actually might be a single product. Further, looking deeper into that list, a person can realize that it is not an easy task to estimate if the products are the same since their descriptions, images, titles, prices may be too different.

Buyers need to get an excellent user experience while searching for products and making purchases on the e-marketplaces. That is why to stimulate purchases, e-commerce platforms are constantly improving the efficiency of collaborating filtering algorithms that form the list of recommended items. One of the steps in this process is items matching. Many e-commerce platforms such as Walmart check if the item already exists in the catalogue before setting up a new one [1]. Other platforms implement items matching step during the search itself. In both cases, this step is necessary and requires intelligent procedures to be fulfilled.

Items matching is usually based on a comparison of textual data provided in item descriptions. Firstly, this data is often incomplete or missed. Secondly, it is represented in a way that puts a strain on matching algorithms. Item's attributes may be called differently and may contain un-normalized values that make the analysis more difficult. And sometimes the description of an item does not contribute much to the buyer's perception. For example, such products as mobile phones, TV-sets, air humidifiers can be precisely described and identified via their technical parameters. On the other hand, clothes, shoes, bags are the products that do not have many attributes for description and even if they do, the show and layout of a thing play a bigger role for item's acquisition. In this situation, items' images should play a major role in matching.

The given paper represents the idea of images matching combined with matching textual descriptions of items. The developed pipeline for data processing allows to collect similar items based on expert judgments, extract their tags and build the core of key attributes that may be used for further matching of new items.

The rest of the paper is organized in the following way. Section 2 considers existing works in the area of images matching in general and in the e-commerce domain in particular. Section 3 describes the scheme of items processing and the tools used for collection of the initial set of items. The experiments with two datasets are given in Section 4. The analysis of results and conclusions are given in Sections 5 and 6.

2 Related Works

Images matching is a complex problem discussed by many researchers in the context of computer vision and augmented reality applications, 3D modeling, visual search, etc. [2]. Many studies represent images as a feature vector and use neural networks for calculating the similarity between images. For example, in [3] authors suggest a Siamese network architecture for generic image retrieval and comparison. The developed architecture showed the best results if it was used together with a pre-trained convolution neural networks obtained during solving the similar problem. Another implementation of the multi-scale Siamese network for image similarity evaluation is given in [4]. Authors incorporated a method of Curriculum learning for online pair mining strategy identification which allowed to provide the increasing difficulty of image pairs during the network training process.

In addition to feature-based methods of image matching that are implemented usually as neural networks, there is also a group of geometric point matching methods. Researchers in [5, 6] describe affine transformation used for images matching. One of the issues resolved by the authors is forming a minimal discrete set of affine transformations applied to each image before matching. Affine transformations together with genetic algorithm implementation for 2D images matching is presented in [6].

Treating an image not only as a set of pixels but as an object that has its internal sense has given a start to the direction of images semantic analysis. For example, authors of [7] worked in the field of semantic segmentation in the computer vision. They represented the 3D-model approach to automatic generation of synthetic images that can be used together with real-world images for semantic segmentation.

The field of e-commerce has brought new statements of images matching problem. Images matching is a non-trivial problem and although everyone realizes that it may improve the efficiency of recommender systems, still researchers find weak points in similarity measures of images and keep on working on new methods. The main goal of images matching in e-marketplace applications is de-duplication of item records. Many studies are aimed at finding identical images that may indicate identical products [1, 8]. In [1] it is stated that due to a large amount of items in the e-marketplace, it is hard to collect manual judgments and it is suggested to use neural networks as feature extractor and cosine similarity. Author of [8] describes the results of using GrabCut computer vision background removal tool for analysis of mobile photo images of new goods that arrive to the fashion store. For image segmentation, Tiramisu DenseNets was used which is a type of convolution architecture. All this was done for checking if the given product already exists in the online store. A kind of similar application of e-commerce images matching is search of all items that look similar to what a buyer has taken a photo of in a real shop but for some reason wants to find online [9].

Generally speaking, images matching is a part of the bigger e-commerce items matching problem. Many researchers suggest different ways to find similar items in the e-stores [1, 10, 11]. For example, in [10] it is proposed to analyze user's browsing of items and build a weighted graph where nodes are the items and edges are the associative connectors that reflect whether two items were watched by the same client during the search. Thus, similar products are identified that can be used for personalized recommendations. However, this study does not analyze item images.

Our previous works [11, 12] represent the approach to grouping similar items based on clustering techniques. Two experiments described deal with mobile phones and bicycles data collected from e-marketplaces by the developed web crawling tool. We can conclude that k-means clustering based on retrieved textual/numerical values of items' attributes gives results that are accurate enough. However, further experiments have shown that items that do not have many attributes in their description cannot be grouped with the help of the proposed method. Images and not attributes have much more sense in the case of such items as clothes or textile objects. That is why the given work is setting sights on the analysis of item images in order to find similar products that can be recommended to a buyer.

3 Collection and Processing of Item Images and Descriptions

The main goal of our research is to find ways to improve the quality of the search for product offers on electronic trading platforms. We have suggested that the solution to this problem should be based on the structuring of item offers. From our point of view, the most natural way is to group similar items. The analysis shows that existing approaches, such as the use of filters or recommendations, are not always effective. The use of filters may limit the choice of products. For example, the product description does not contain information about all color options, the price includes shipping costs and does not match the selected range, an error may be made in the model name, the brand is not specified, etc. The results of the recommendation systems vary on different trading platforms and often use mixed algorithms that take into account not only the characteristics of goods but also the behavior of users. As a result, some products may not fall into the search selection or similarity group. On the other hand, we noticed that a person surfing via an e-marketplace always identifies similar products. In this case, the person relies on the image and a brief description of the item. The main idea of this study is to create a core of keywords that determine the description of a similarity group.

Analyzing a lot of up-to-date evaluation approaches, we found out the pairwise comparison is the best way for humans. Following the Theory of Intelligence [13], we propose to take into consideration the ability of human intelligence to perceive images. In the process of examining human behavior, it is important to research internal subjective state and information processing which cause a particular way of behavior. Examples of internal states are images and representations of real goods. We present the image of the product and get the perception as an assessment of the internal state. We suggest that the perception of images of similar products is the same. A person compares images of goods. If the goods are similar, then the answer should be "yes".

According to [13], the internal relations can be represented with a predicate function named predicate of equivalence. The predicate of equivalence is a two-place predicate, which is reflective, symmetrical, and transitive. By analogy with the predicate of equivalence, we propose to consider the predicate of goods similarity. The reflexiveness of the predicate of similarity means that similar items are represented by similar images. The symmetry of the predicate of similarity means that objects remain similar even if you swap places of their images. The transitiveness of the predicate of similarity means that if the first and the second items are similar and the second and the third items are similar too, then the first and the third items are also similar.

Therefore, just monitoring how people implement items matching we can obtain data about items similarity. We suggest applying pairwise comparison of item images to obtain similarity estimation. Based on the processing of the pairwise comparison matrix we can determine the groups of similar items. We make a hypothesis that the group of similar items has the common core description, that is why we suggest building the core of the keywords for each group. So, we can extend the group of similar items based on comparing the new item with the core of keywords. The general scheme of processing of items is presented in fig. 1.

Collecting images and descriptions is performed with the help of grabbing and parsing software. The raw data sets are inspected manually. We have developed our own mobile application for pairwise comparison of items. We suggest using a crowdsourcing approach in order to collect estimations of item matches. Due to this reason, we have asked volunteers to upload our mobile application and do comparisons. The developed software component is an Android-mobile client.

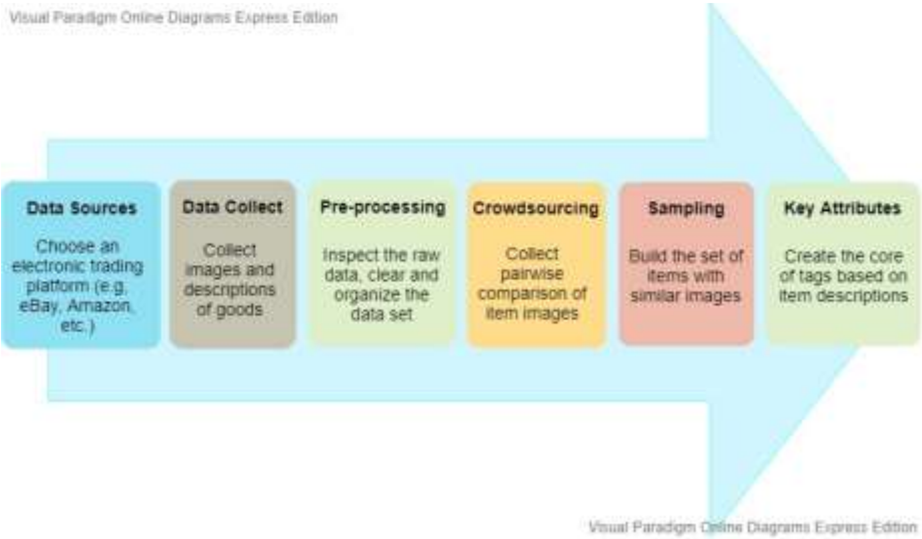

Fig. 1. The general scheme of processing of items

The user can compare images in pairs and indicate the similarity or difference of items offered for comparison. The data provided by the user is collected in cloud storage and further can be used for the next steps of processing. The result of the comparison is sent directly to Firebase Cloud node via HTTP-protocol where it is validated and stored. Each user session has a unique ID to make data analysis more efficient. If the user wants to finish the comparing, the "Finish" button should be pressed and the application would be closed. The image comparison screen is given in fig. 2.

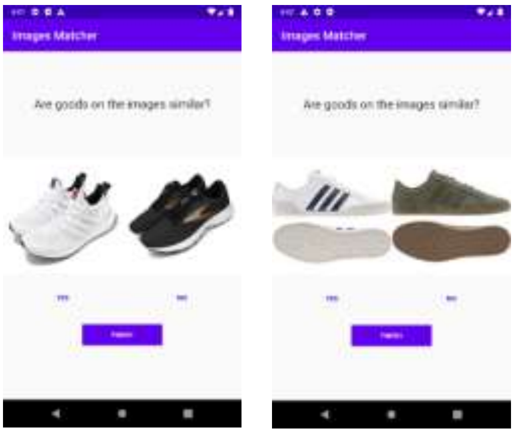

Fig. 2. The image comparison screen (examples)

Therefore, we collect images and descriptions of items, then estimate the similarity based on image matching, and finally process item description to build the core of tags for each group of similar items. The given study is focused on developing the key attributes of similar items based on the item's image tag comparison.

4 Experiments

4.1 Data Set Collection and Preparation

In this paper, we have checked the idea that if we know item groups with similar objects, then in the future we can easily find the correct group for new items. Thus, the main idea to make an experiment is checking the assumptions: 1) that similar items have the same core of tags; 2) that group of a new object can be easily found if the core for a group of objects is known. In the first step, we create the dataset. Each item in our dataset has an image and a text description. The example of such an item is presented in Table 1.

Table 1. An example of item in dataset.

Image	Tags
	A brand-new, Unused, Unworn, Sneakers, Breathable, Light-weight, Comfort, G27706, Does not apply, Full Year Article, Leather, adidas, Fabric, Does not apply, 2010-2019, Low Top, Leisure, Rubberband, Unisex, Lacing, Casual, adidas Continental, adidas Originals, adidas Continental 80 , Standard, Casual Shoes, White / Scarlet / Collegiate Navy, Fitness Studio & Training

For our experiments, we have transformed the item description into a set of tags. We prepared two datasets according to our data processing pipeline. The first dataset is the set of similar items gathered from the eBay trading platform (https://www.ebay.com/). We chose the data about white sneakers. The second dataset contains different items, for example, different types of shoes, bags, etc. from the same source. Table 2 shows information about datasets.

Table 2. Information about test data sets.

	Number of items	Number of tags
Dataset_1	79	2081
Dataset_2	165	2382
Total	244	4463

After the analysis of our datasets, we decided to clean the data, because tags for all items vary a lot. For example, we have such tags as "M" or "9.5" as the size of shoes or "2010-2019", "B42000", "light-weight", "Running & Jogging". A lot of tags contain different punctuation such as "&", "-", ",", ")", "(", etc. All stages of the cleaning process are presented in Fig. 3.

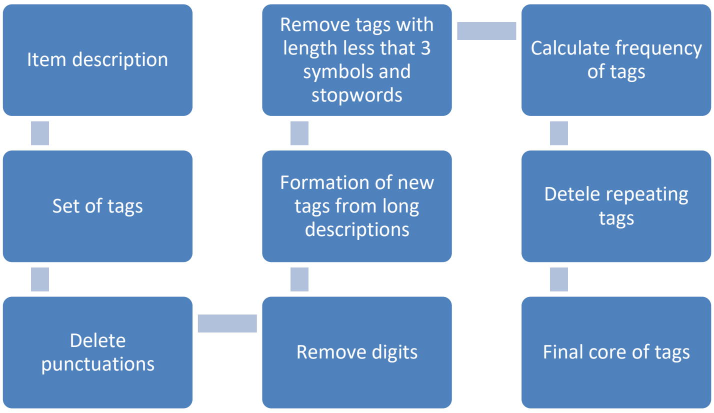

Fig. 3. Stages of the cleaning process

In addition to removing punctuation and other characters, we needed to convert long tags (such as "Running & Jogging") to a set of short tags ("Running" and "Jogging"). For example, "White / Scarlet / Collegiate Navy" was transformed into "White", "Scarlet", "Collegiate" and "Navy". We decided to remove stopwords and tags with length less than 3 symbols, these tags are not informative for the dataset. In order to get the tag core, we also need to calculate the tag frequency for the collection and remove duplicate tags. The number of tags after each stage is shown in Fig. 4.

4.2 Experiments with Similarity Tags

Having the cleaned tag list sorted by descending tag frequency, we created the tag core for the Dataset_1. The algorithm of tag core creation was described in detail in work [14].

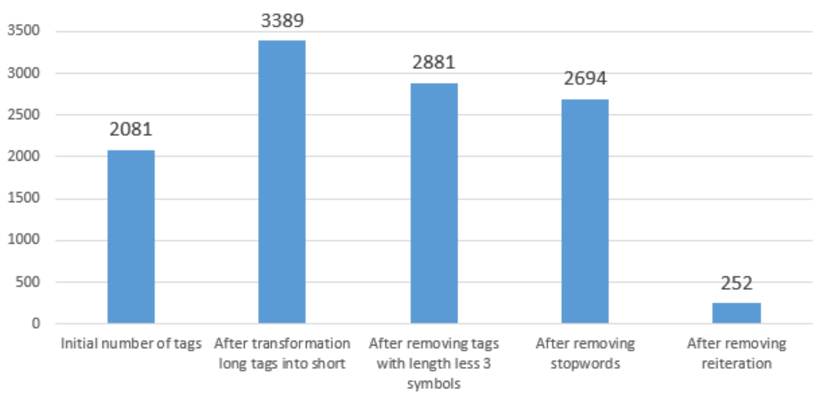

Fig. 4. The number of tags of Dataset_1

This algorithm is based on word2vec model. The proposed algorithm is not complicated, but it takes into account the semantic similarity of words using word2vec and the tag frequency after the pre-processing stage. The results of these experiments are presented in Table. 3.

Table 3. The core tags (for similarity value > 0.75).

	Tag core
Dataset_1	leather, white, comfort, low, shoes, fitness, summer, trainers, toe, sports, basketball, school, outsole, skate

We took a similarity value over 0.75. It was selected based on an analysis of the tags received as well as their number. As Table 3 shows, we received a fairly short list of tags when the value of the similarity measure is more than 0.75. As a test of how correctly our tag core describes our data set, we compared each item from Dataset_1 with our core using function similarity from Spacy library (Fig. 5). The mean of our results is 0.87, min is 0.73 and max is 0.92. These results show that we have a good core for our Dataset_1 and this tag core describes our collection of items pretty well.

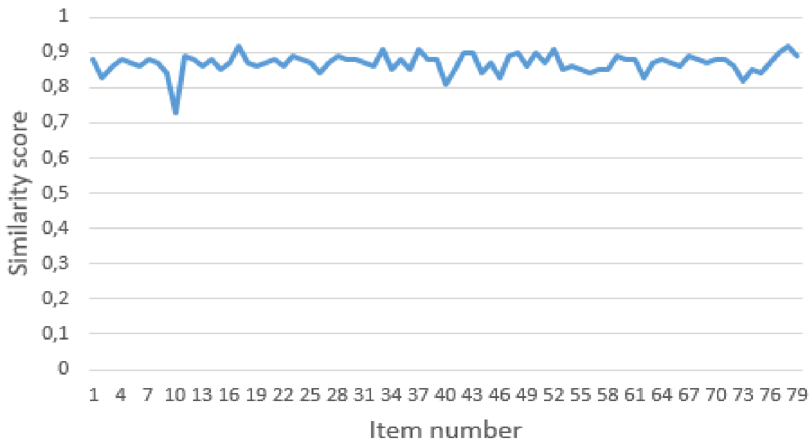

Fig. 5. The value of similarity metric between items of Dataset_1 and the tag core

On the next step, we made the experiments with the tag core and Dataset_2. In this case, we compare each item's tags with our core and receive a similarity score. This score helps us to answer the question "Does this item belong to this group or not?". The Fig. 6 shows the similarity values for this experiment. The mean of our results is 0.74, min is 0.34 and max is 0.85. These results show that we have very different items in Dataset_2 - both very similar and dissimilar. We analyzed these results and prepared some samples. These samples are presented in Table 4. The examples from Table 4 demonstrate that our approach works correctly with different types of items.

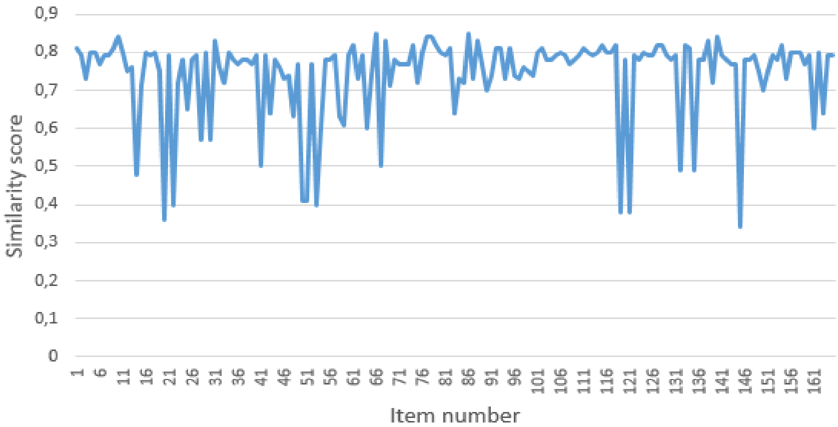

Fig. 6. The value of the similarity metric between items of Dataset_2 and tag core

Table 4. Work examples.

Image	Tag core	Similarity score
	*A brand-new, Unused, Unworn , 100% Authentic Guaranteed , Fw5422 , Us Size , Medium , Running, Cross Training , Men , Running, Cross Training , Synthetic , White , Ultraboost Hk , Adidas , Athletic , Synthetic*	0,8
	*A brand-new, Unused, Unworn , 214022 , Not specified , Womens , Not specified , Asics , Regular*	0,6
	*A brand-new, Unused, Unworn , 100% Authentic Guaranteed , Cm997hebd , Us Size , Medium , Running, Cross Training , Men , Running, Cross Training , Synthetic , Blue , Cm997heb D , New Balance , Athletic , Synthetic*	0,63
	*A brand-new, Unused, Unworn , Gummi , Schnürsenkel , Atmungsaktiv , Turnschuhe , Fitness, Fitnessstudio & Training, Spazieren , Schwarz , Straße , Flex 2019 RN , dunkel , Straße , Phylon-Mittelsohle von Nike (EVA) , Synthetik , Nike , Flex , Synthetik , Vietnam , Textil , AQ7487-005 , Synthetik*	0,57

5 Evaluation and Analysis of Results

We evaluate the results of the experiment in two stages. The results presented above show that the constructed core describes the dataset well. Inside the dataset items are similar, but not always the same products. The analysis showed that the obtained estimates for identical pictures are equal. In these cases, the sets of tags also match. This confirms the assumption that the grouping can be carried out on the basis of assessing the similarity of the product description and core of tags for the group. It should be noted that for identical items with different images, we find differences in similarity values and in the sets of tags respectively. This may be because different sellers provide such descriptions.

The main idea of building a core of tags is to be able to determine whether a new product will fall into a group of similar goods. In order to estimate the proposed way to matching items, we use the Dataset_2. It contains 162 items of sneakers, shoes, and some other goods. According to the data processing pipeline discussed above, the sets of tags for each item description are created. We evaluate manually all the items whether they match items from the Dataset_1. Then we compare manual estimates with the similarity score of each item. We use the 0.75 similarity score as an indicator to join the item group. As a result, we obtain *Total Items* = 162, *True Positive Items* = 101, *True Negative Items* = 19, *False Positive Items* = 14, *False Negative Items* = 28. Therefore, *Accuracy* = 0.74, *Precision* = 0.88, *Recall* = 0.89. Manually classified items are all sneakers with a similar shape or similar models without taking into account their color. Thus, we can conclude that the proposed approach is validated, but it should be improved by studying the similarity score more deeply.

6 Conclusions and Future Work

The given work represents the experimental pipeline of the core of key attributes construction for e-commerce items. It is suggested to build such a core from the tags of the items that have been acclaimed as similar by the experts based on images comparison. The method proposed was proved to be suitable for constructing the core for sneakers. The two open questions in method application are: 1) the definition of the threshold for attributes inclusion to the core since different values lead to a different number of keywords included; 2) the adaptation of the threshold for the analysis of new items that have to be compared with the core.

The future directions of this research include testing the method on other types of products, not clothes (shoes) but household appliances, for example. Also, the application of the method for bigger initial samples (including millions of items) obviously will require its modification and additional testing. And finally, the evaluation of search and filtering algorithms performance with and without application of the keywords core is planned for future work.

7 References

1. More, A.: Product Matching in E-commerce Using Deep Learning (2017). https://medium.com/walmartlabs/product-matching-in-ecommerce-4f19b6aebaca
2. Walia, E., Suneja, A.: A Conceptual Study on Image Matching Techniques. Global Journal of Computer Science and Technology, 10 (12), 83-88 (2010).
3. Melekhov, I., Kannala, J., Rahtu, E.: Siamese network features for image matching. Conference: 2016 23rd International Conference on Pattern Recognition (ICPR) pp. 378-383 (2016). DOI: 10.1109/ICPR.2016.7899663.
4. Appalaraju, S., Chaoji, V.: Image similarity using Deep CNN and Curriculum Learning (2017).
5. Rodríguez, M., Delon, J., Morel, J.-M.: Fast Affine Invariant Image Matching. Image Processing On Line. 8, 251-281 (2018). https://doi.org/10.5201/ipol.2018.225.
6. Bazargani, M., Anjos, A., Lobo, F., Mollahosseini, A., Shahbazkia, H.: Affine Image Registration Transformation Estimation Using a Real Coded Genetic Algorithm with SBX (2012). doi.org/10.1145/2330784.2330990.
7. Zhang, Y., Wu, Z., Zhou, Z., Wang, Y.: Synthesizing Training Images for Semantic Segmentation. In: Wang Y., Jiang Z., Peng Y. (eds) Image and Graphics Technologies and Applications. IGTA 2018. Communications in Computer and Information Science, vol 875, pp. 220-227. Springer, Singapore (2018). https://doi.org/10.1007/978-981-13-1702-6_22
8. Yam, C.-Y.: Deep Learning Image Segmentation for Ecommerce Catalogue Visual Search (2018). https://devblogs.microsoft.com/cse/2018/04/18/deep-learning-image-segmentation-for-ecommerce-catalogue-visual-search/
9. Kiapour, M. H., Han, X., Lazebnik, S., Berg, A. C., Berg, T. L.: Where to Buy It: Matching Street Clothing Photos in Online Shops. IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3343-3351. Santiago (2015).
10. Madvariya, A., Borar, S.: Discovering Similar Products in Fashion E-commerce. eCOM@SIGIR. (2017).
11. Cherednichenko, O., Vovk, M., Kanishcheva, O., Godlevskyi M.: Towards Improving the Search Quality on the Trading Platforms. In: S.Wrycza, J. Maslankowski(Eds): 11th SIGSAND/PLAIS 2018, LNBIP 333 pp. 21-30. Springer (2018).
12. Cherednichenko, O., Vovk, M., Kanishcheva, O., Godlevskyi M.: Studying Items Similarity for Dependable Buying on Electronic Marketplaces. Proc. 2nd Int. Conf. On Computational Linguistics and Intelligent Systems (COLINS), Volume I: Main Conference (Lviv, Ukraine, 2018). Vol. 2136. pp.78-89 CEUR-WS (2018).
13. Bondarenko, M. F., Shabanov-Kushnarenko, U. P.: Theory of intelligence: a Handbook. SMIT Company, Kharkiv (2006).
14. Kanishcheva, O., Cherednichenko, O., Sharonova, N.: Image Tag Core Generation. In: 1st International Workshop on Digital Content & Smart Multimedia (DCSMart 2019) Ukraine, CEUR Workshop Proceedings, Volume 1, pp. 35-44. Lviv (2019). http://ceur-ws.org/Vol-2533/preface.pdf