**Filtering Instagram Hashtags Through Crowdtagging and the HITS Algorithm**

**Abstract:**

Instagram is a rich source for mining descriptive tags for images and multimedia in general. The tags–image pairs can be used to train automatic image annotation (AIA) systems in accordance with the learning by example paradigm. In previous studies, we had concluded that, on average, 20% of the Instagram hashtags are related to the actual visual content of the image they accompany, i.e., they are descriptive hashtags, while there are many irrelevant hashtags, i.e., stop-hashtags, that are used across totally different images just for gathering clicks and for searchability enhancement. In this paper, we present a novel methodology, based on the principles of collective intelligence that helps in locating those hashtags. In particular, we show that the application of a modified version of the well-known hyperlinkinduced topic search (HITS) algorithm, in a crowdtagging context, provides an effective and consistent way for finding pairs of Instagram images and hashtags, which lead to representative and noise-free training sets for content-based image retrieval. As a proof of concept, we used the crowdsourcing platform *Figure-eight* to allow collective intelligence to be gathered in the form of tag selection (crowdtagging) for Instagram hashtags. The crowdtagging data of *Figure-eight* are used to form bipartite graphs in which the first type of nodes corresponds to the annotators and the second type to the hashtags they selected. The HITS algorithm is first used to rank the annotators in terms of their effectiveness in the crowdtagging task and then to identify the right hashtags per image.

**Existing System:**

Instagram is a rich source for mining descriptive tags for images and multimedia in general. The tags–image pairs can be used to train automatic image annotation (AIA) systems in accordance with the learning by example paradigm. In previous studies, we had concluded that, on average, 20% of the Instagram hashtags are related to the actual visual content of the image they accompany, i.e., they are descriptive hashtags, while there are many irrelevant hashtags, i.e., stop-hashtags, that are used across totally different images just for gathering clicks and for searchability enhancement.

**Proposed System:**

we present a novel methodology, based on the principles of collective intelligence that helps in locating those hashtags. In particular, we show that the application of a modified version of the well-known hyperlinkinduced topic search (HITS) algorithm, in a crowdtagging context, provides an effective and consistent way for finding pairs of Instagram images and hashtags, which lead to representative and noise-free training sets for content-based image retrieval. As a proof of concept, we used the crowdsourcing platform *Figure-eight* to allow collective intelligence to be gathered in the form of tag selection (crowdtagging) for Instagram hashtags. The crowdtagging data of *Figure-eight* are used to form bipartite graphs in which the first type of nodes corresponds to the annotators and the second type to the hashtags they selected. The HITS algorithm is first used to rank the annotators in terms of their effectiveness in the crowdtagging task and then to identify the right hashtags per image.