"Image Search IR System"

• • •

WS 2020-21 Information Retrieval Project

Claudio Maggioni

Objectives

- Build a text retrieval system that allows for searching images
- The system should support result clustering
- Results must be able to be filtered based on the cluster they belong

Crawlers

- Sites crawled are *Shutterstock, Flickr* and *123RF.*
- Two are stock image sites, one is a user image publishing website
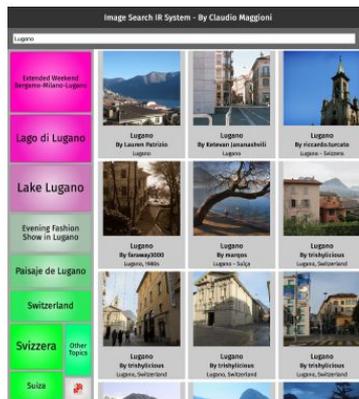- For *Flickr,* Splash, a headless browser extension to Scrapy, has been used

Solr

- Solr configuration was inspired by the *Techproducts*
- Clustering is supporting using *Carrot2* technology
- As search field title, description and author of the image are used

User Interface

(a) The UI, when opened, prompts to insert a query in the input field and press Enter. Here the user typed "Lugano".

(b) After the user inputs a query and presses Enter, resulting images are shown on the right. Found clusters are shown on the left using FoamTree.

(c) When a user clicks a cluster, results are filtered depending on the cluster clicked. If the user clicks again on the cluster, filtering is removed.

User Evaluation

- Three test subjects, all 3rd year bachelor students at USI
- Generally test subjects are satisfied with the product
- Two improvements emerged
  - Add search box submit button
  - Make clustering feature clearer

Demo

Questions