《信息检索 》第一次作业

计83李天勤2018080106 George Li

作业：在网上寻找任意类型的、与信息检索有关的任意可运行系统（针对text、image、video、audio等均可。商用系统及尚处于研究中的demo系统都可以。我们鼓励大家尝试花样尽可能多的“搜索引擎”（如可以包括情感分类、OpenQA等），查询query可以不限于文字，例如可以是一张图片），键入一定量的、能够说明问题的相关query，测试其功能，对结果给出某种形式的定量评估，并根据自己的理解作相应分析。(交报告。适当的时候，将由助教组织课堂研讨）。

Search Engine: Bing

Microsoft Bing Visual Search <https://www.bing.com/visualsearch?FORM=ILPVIS>

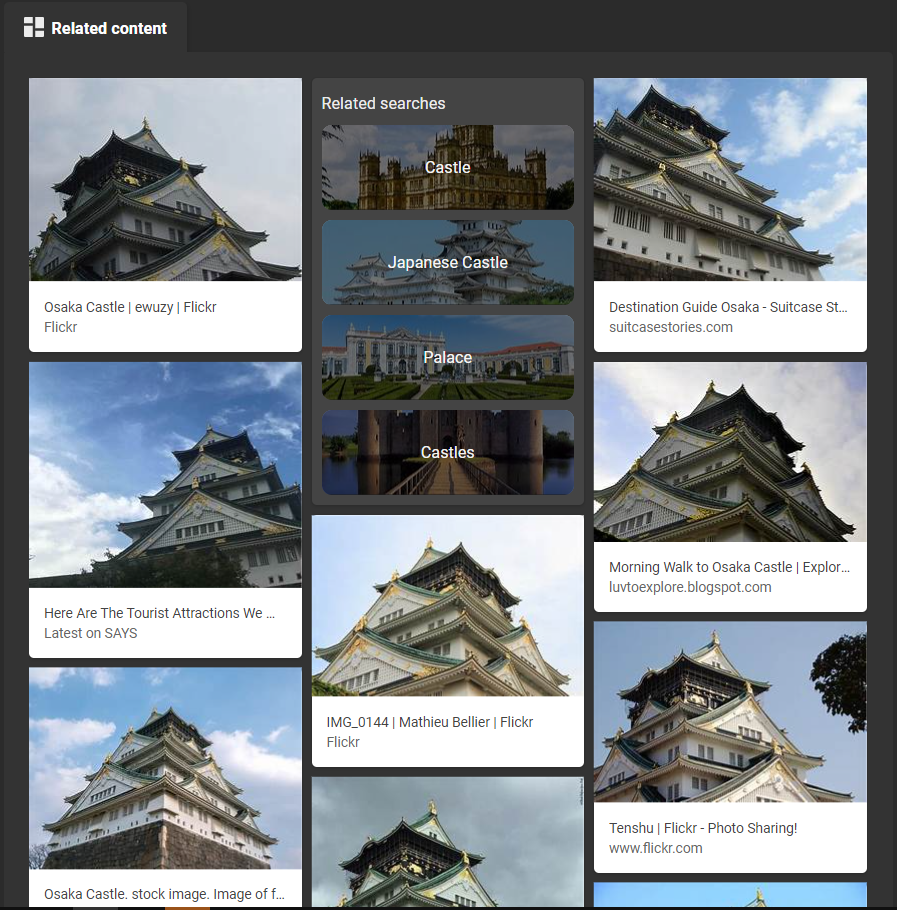
Query: Image

Return: Related Images

The query image that I used was a picture that I had taken in a trip to Osaka Japan, it is the Osaka Castle



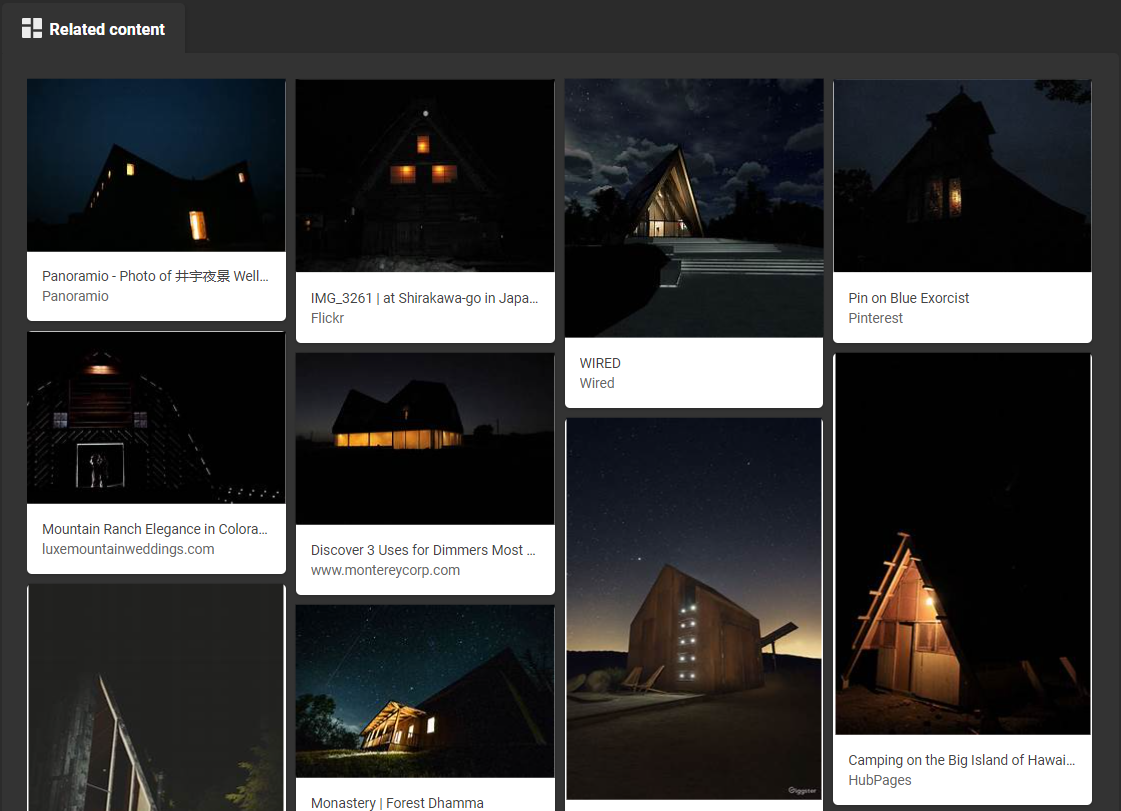
These are the results



All of the results are images of the Osaka Castle. It even returns related searches that have a connection with Osaka Castle, and architecture in general.

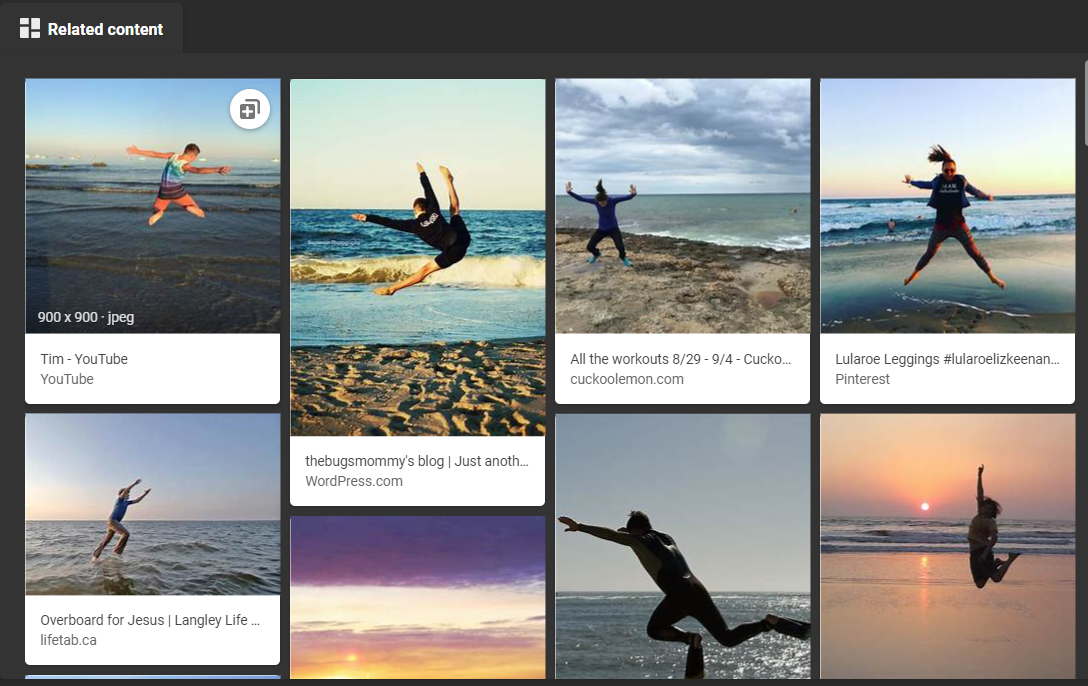
The second image that I searched on Bing is a picture of 阿那亚礼堂 at 秦皇岛。





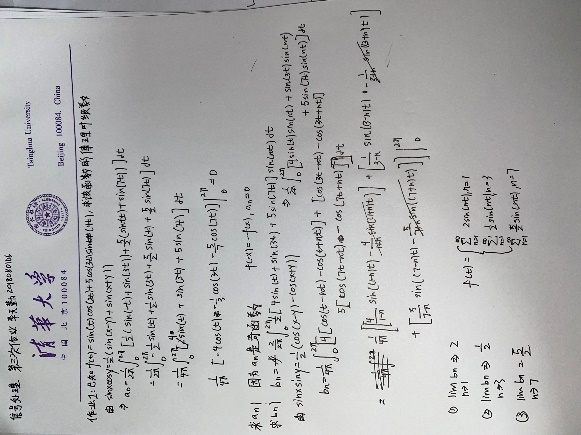
From the results, we can see that the images do not match the one that is submitted. However, the images are quite similar in context, composition, and color. This may be due to the lack of entries of this image since it is a landmark in China, which Bing lacks.

Next is an image of a person on the beach (it is me).

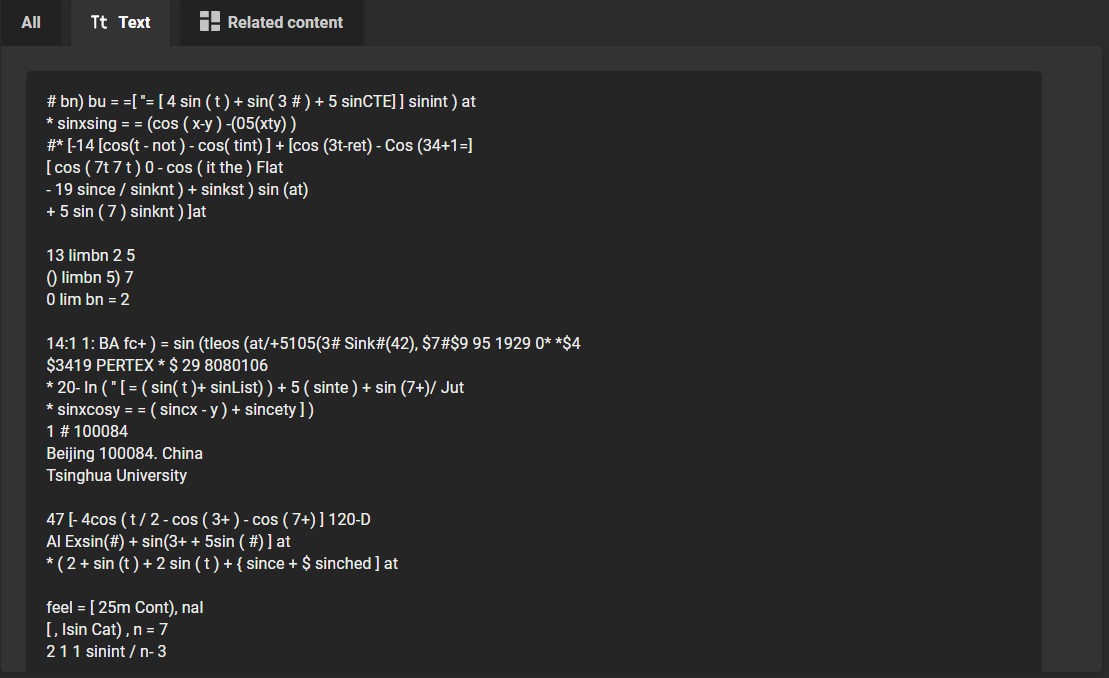
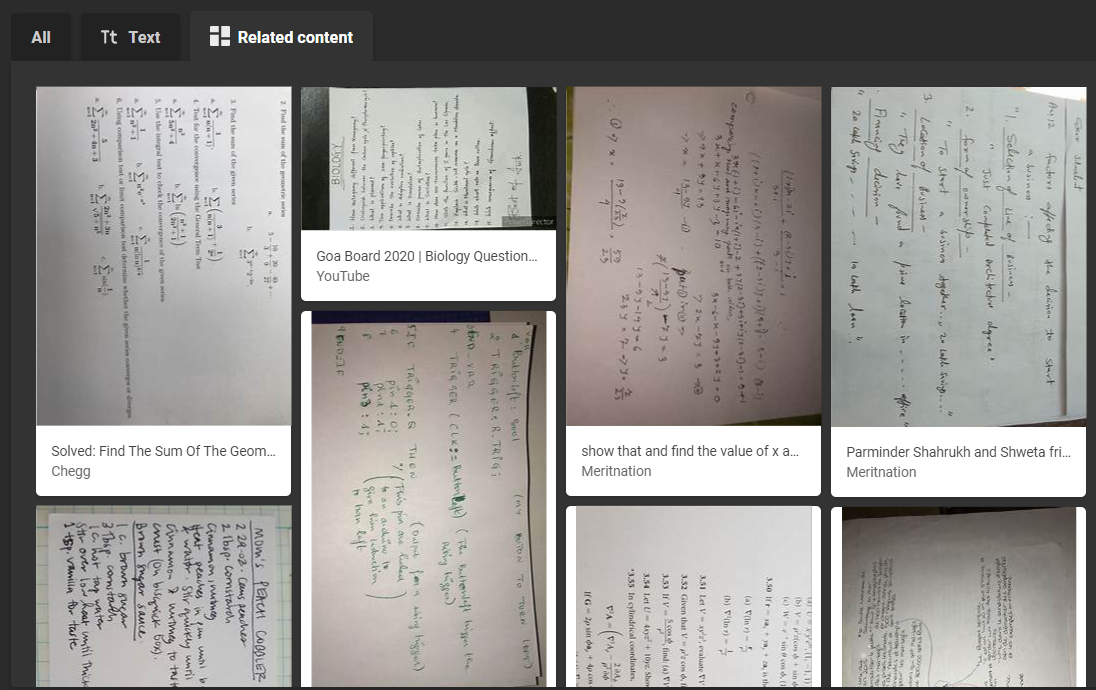


The results contain images similar to the query, containing people at beaches.

Next is an image of my homework.



We can see from the results that the search engine recognizes text in an image. It can also extract text from the image, and detect words like “100084, Beijing 100084. China Tsinghua University.”



The second test is with 文字. I will be using 国内版。

Query: 蛋白粉vs蛋白粉的正确使用方法

第一个query：蛋白粉

第二个query: 蛋白粉的正确使用方法

Result: From the results, we can generalize two things. If the input is a word, the content in the output link generally directly contains the word, on the other hand, if the input is a sentence, the sentence will be segmented and then matched, and the output link generally contains the segmented words.