

# Image Retrieval Using Image Captioning

Safaa Y. Diab

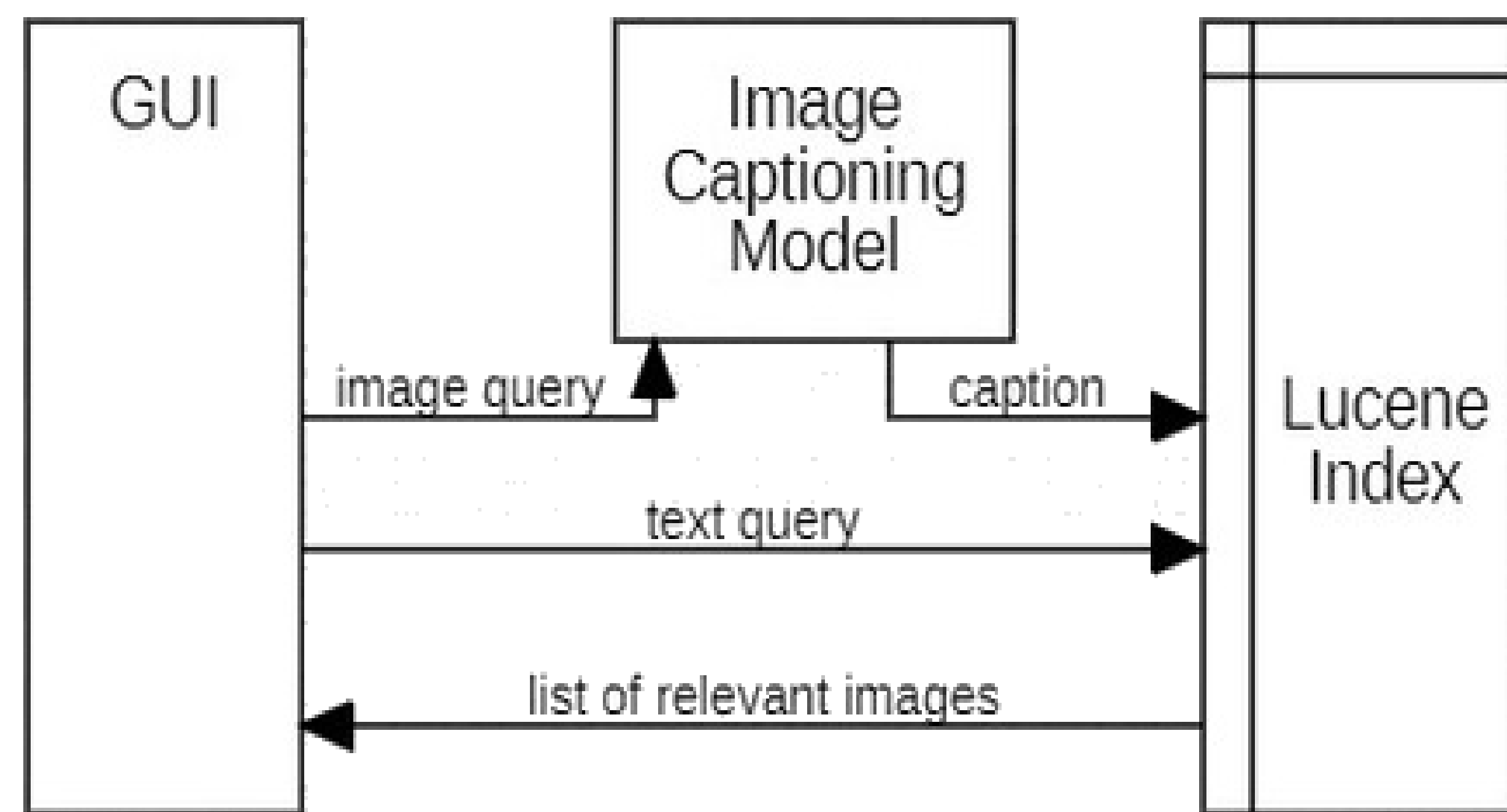
## Objectives:

- Retrieve similar images given an image.
- Search for images given a text query.

## Data:

- MSCOCO 2014<sup>1</sup>: a large-scale object detection, segmentation, and captioning dataset.
- 80k images
- 5 captions per image

## How It Works?



## Image Captioning model:<sup>2</sup>

- Show, Attend and Tell: Neural Image Caption Generation with Visual Attention.<sup>3</sup>
- Uses Inception V3.
- Trained on MSCOCO dataset.

## Indexing:

### Text based Indexing:

Given the generated caption or text query. We need to lookup for relevant images by searching their labels. Hence an index is created where:

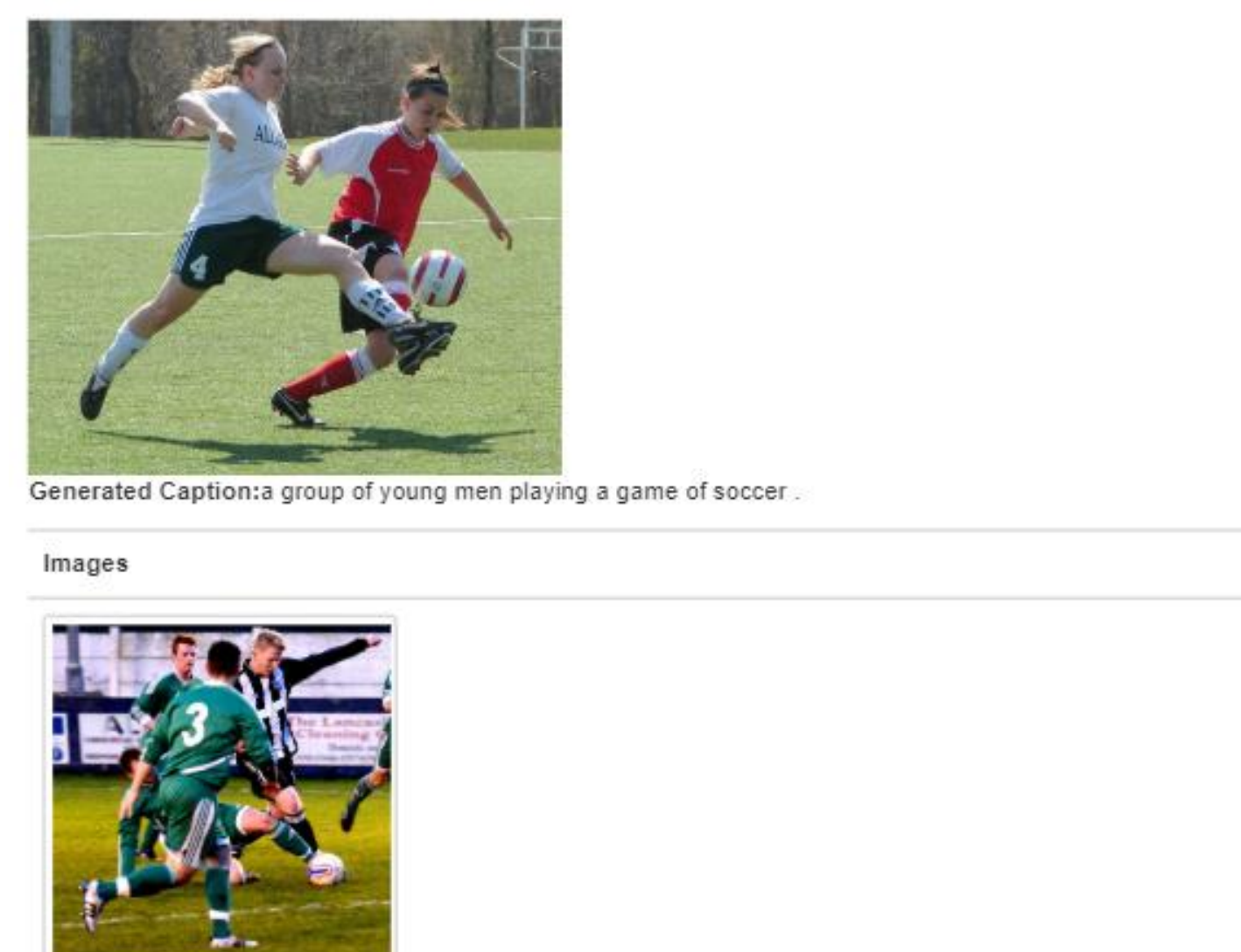
- For each image a doc containing the 5 labels is created.
- Stemming and stop words are removed.
- VSM similarity is used as a scoring model.

### LIRE Image Indexing<sup>4</sup>:

LIRE is a library that creates a Lucene index of image features for content based image retrieval (CBIR). Some of the extracted features used:

- CEDD and FCTH.
- auto color correlogram.
- Visual words based on SIFT and SURF

## Example:

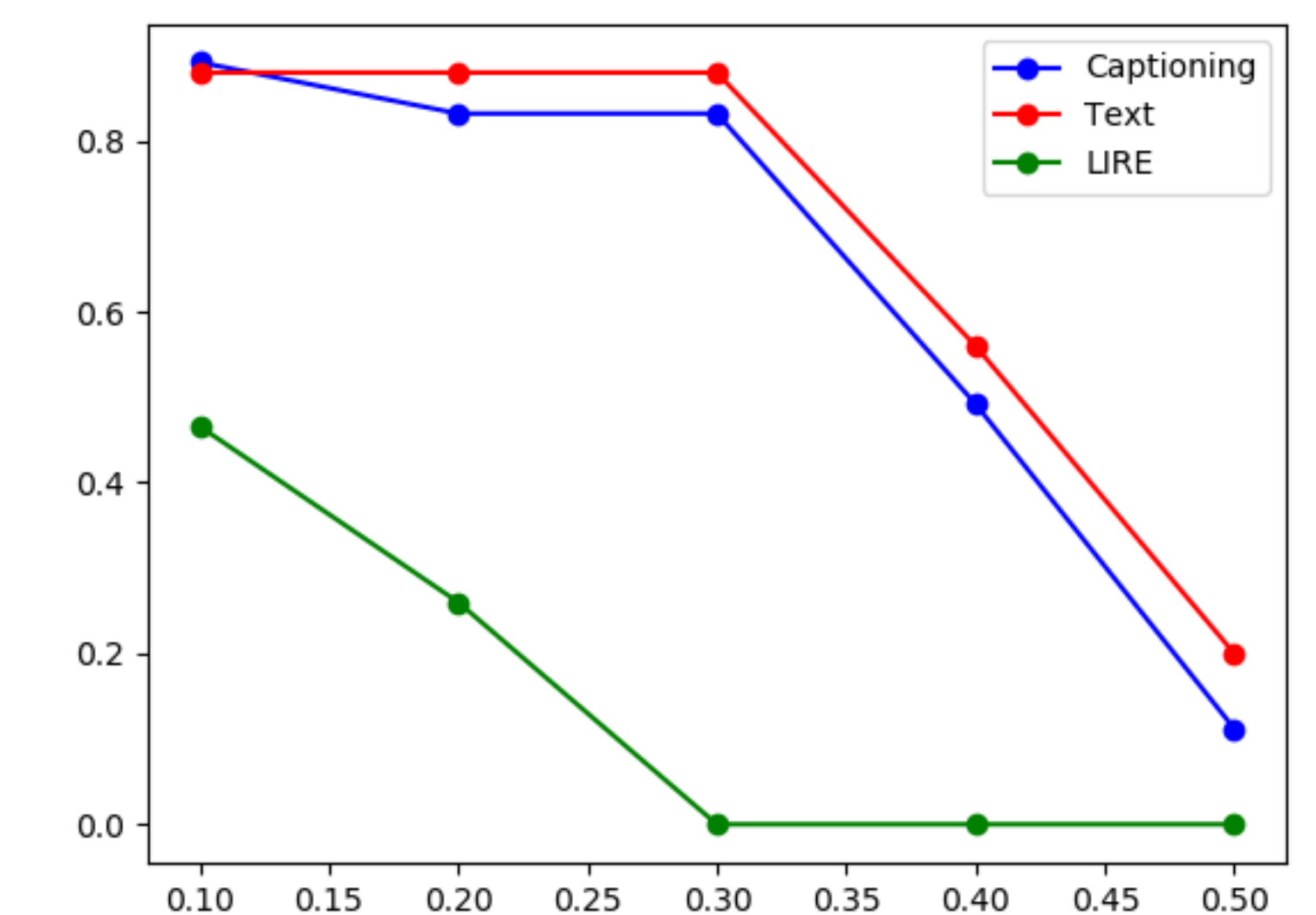


## Evaluation & Results:

A precision-recall curve has been drawn in order to evaluate the performance of the following:

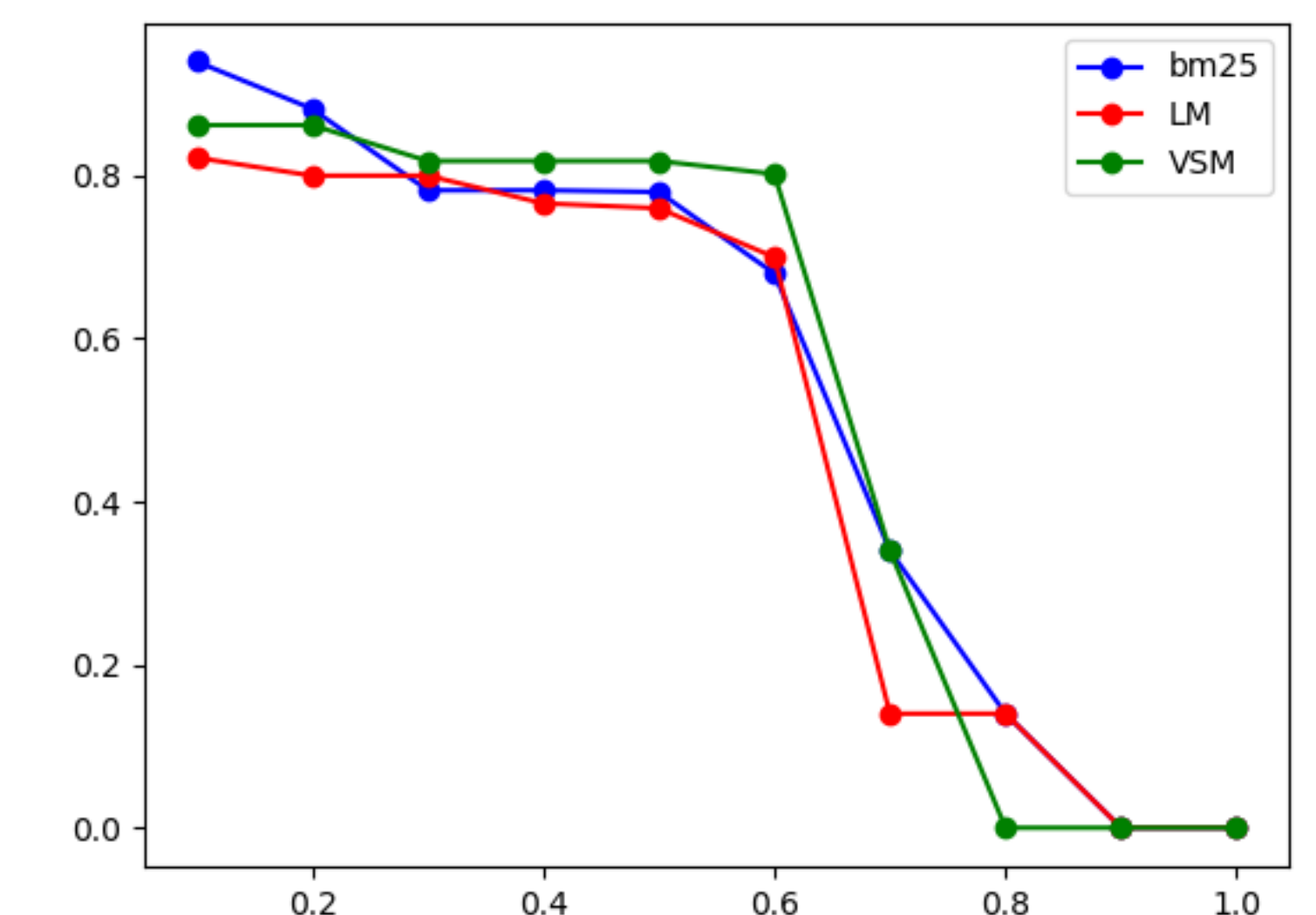
Image Retrieval using:

- Image captioning
- Text query
- LIRE



Different scoring models:

- BM25
- LM
- VSM



	Caption	Text	LIRE
MAP	0.82	0.91	0.46
NDCG	0.84	0.89	0.89

## Technologies:



## References:

- [cocodataset.org/#home](http://cocodataset.org/#home)
- [github.com/Gurpreetsingh9465/imageCaption](https://github.com/Gurpreetsingh9465/imageCaption)
- [Vinyals Show and Tell 2015 CVPR paper.pdf](https://arxiv.org/pdf/1502.03214v1.pdf)
- [www.semanticmetadata.net/wiki/](http://www.semanticmetadata.net/wiki/)