Query Intent Detection from the SEO Perspective

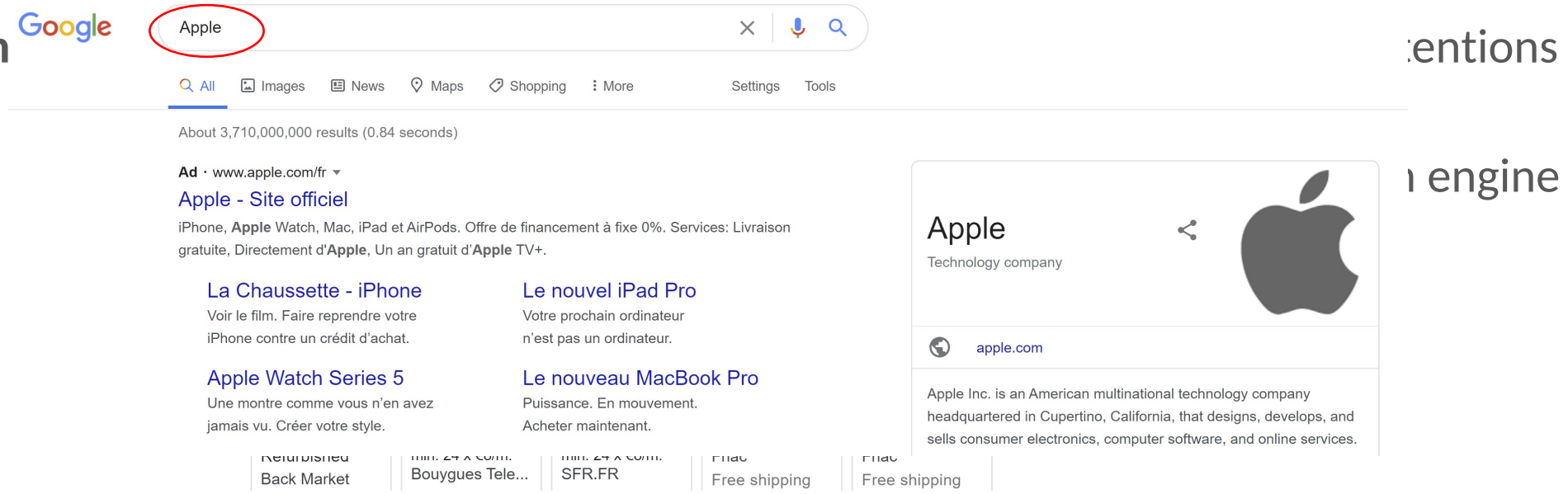
Samin Mohammadi, Mathieu Chapon, Arthur Frémond

Samin.mohammadi@search-foresight.com

August 2020

Introduction

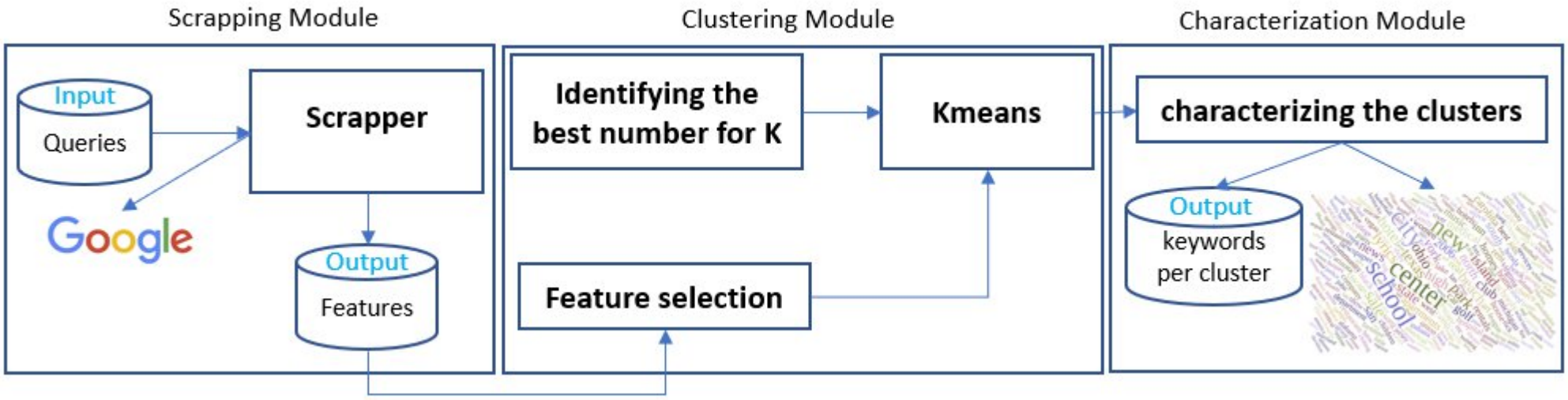
- **SEO:** Search Engine Optimization
  - Increase the visibility
  - Bring the website's rank upper in the search engine result
- **User query's intentions:** Informational, Navigational, Transactional
- **Query inten** :entions
- **Application** engine expects



Intent detection

- **Problem statement:** Identifying the intent of a given user query
  - Input: a set of user queries (short texts)
  - Output: intent of queries
  - Model: Kmeans clustering
  - Challenge: Ending up with intents different from the conventional intents

- **Model architecture:**



Architecture

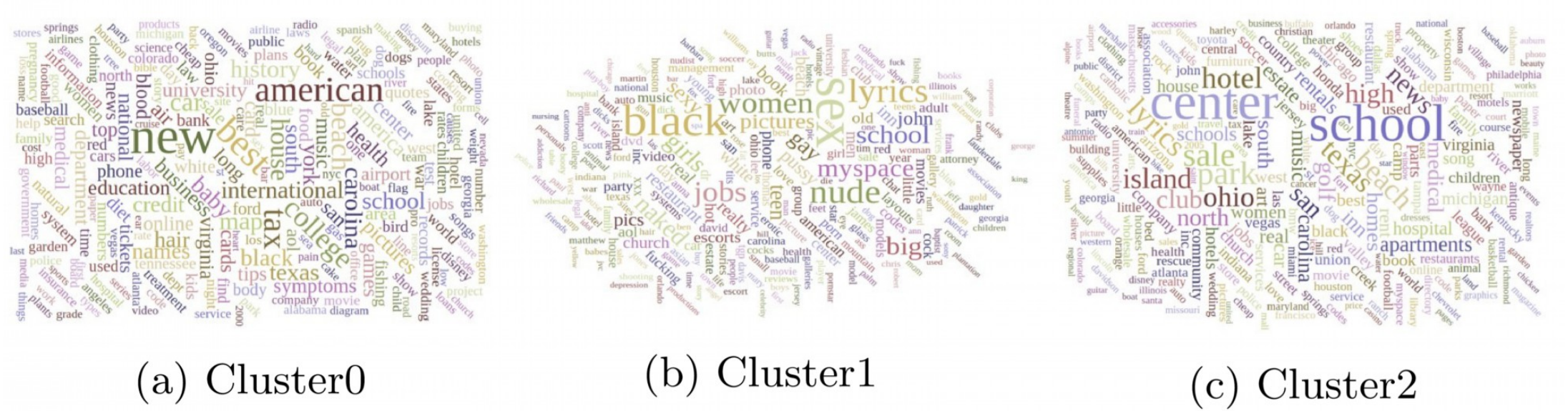
- **Scrapper**
  - Data Scrapper, implemented by Python, sends a query as a request to Google and collects the results provided by Google in the form of HTML.
- **Feature Extraction**
  - In HTML resutls, different types of answers are provided including videos, images, featurd snippet, rich snippet, knowledge graph, direct answer. We considered them as features.
- **Dataset**

Dataset name	#Queries	Manual-Labels		
		Informational	Navigational	Transactional
Selected-AOL	30K	23700	4574	1678

- **Clustering**
  - Based on Elbow method Number of clusters = 3

Intent Detection Result

- **WordCloud Representation**



- **Cluster0:** {New, Best, American} + {Tax, Business, car, Education, College, Uniersity}
  - Qualitative Information
- **Cluster1:** {Black, Sex, Nude, Women}
  - Racist/Sexual Intent
- **Cluster2:** {Center, School, Park, Hotel, Beach, Island}
  - Local/Place Information

Results & Conclusion

- Extract the most frequent words for each cluster and label user queries using them

Clustering vs Vocabulary-based intent tagging result

		Vocabulary-based intents(predicted)				
		Informational	Local/place	Sexual/Racism	Precision	Recall
Clustering intents (actual)	Informational	1232	54	25	0.46%	**0.94%**
	Local/place	904	141	26	**0.64%**	0.13%
	Sexual/Racism	519	25	70	**0.58%**	0.11%

- Conclusion:
  - Proposed a model to identify the intent of user query
    - Scrapes the Google SERPs
    - Uses new features
    - Identifies new intents different than the conventional ones

contact

Samin Mohammadi

Samin.mohammadi@search-foresight.com

+ 33 1 74 18 29 69

Slideshare.net/S4sight

@S4sight

www.search-foresight.com