



Query Intent Detection from the SEO Perspective

Samin Mohammadi, Mathieu Chapon, Arthur Frémond

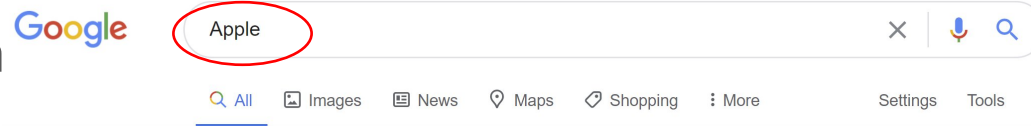
Samin.mohammadi@search-foresight.com

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 intentions**



• **Query intentions**

• **Application expects**

About 3,710,000,000 results (0.84 seconds)

Ad · www.apple.com/fr

[Apple - Site officiel](#)

iPhone, **Apple Watch**, Mac, iPad et AirPods. Offre de financement à fixe 0%. Services: Livraison gratuite, Directement d'**Apple**, Un an gratuit d'**Apple TV+**.

[La Chaussette - iPhone](#)

Voir le film. Faire reprendre votre iPhone contre un crédit d'achat.

[Le nouvel iPad Pro](#)

Votre prochain ordinateur n'est pas un ordinateur.

[Apple Watch Series 5](#)

Une montre comme vous n'en avez jamais vu. Créer votre style.

[Le nouveau MacBook Pro](#)

Puissance. En mouvement. Acheter maintenant.

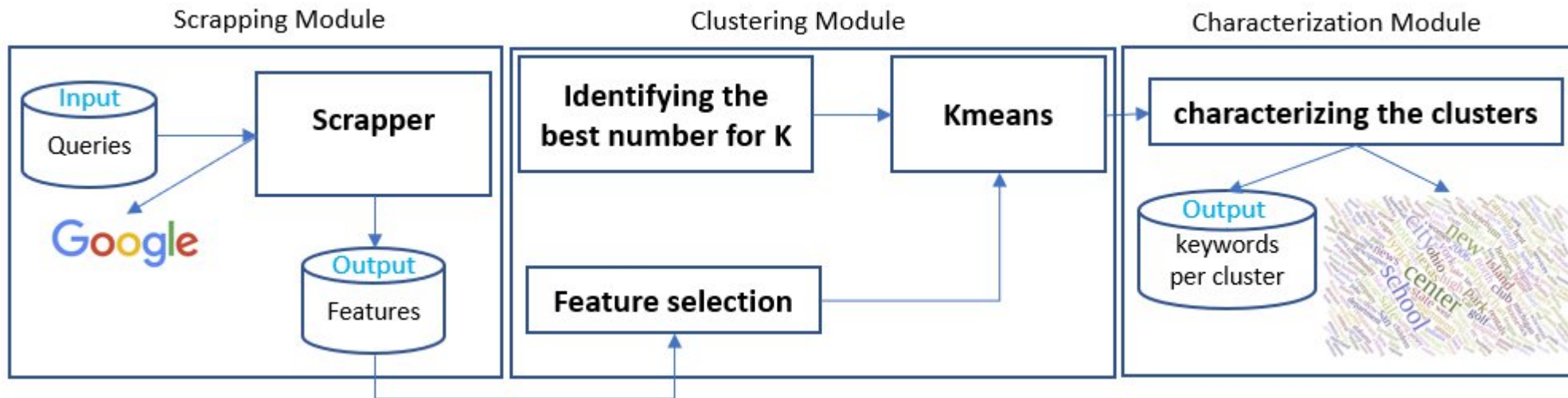
A knowledge panel for Apple. It features the word "Apple" in a large font, followed by "Technology company" in a smaller font. To the right is the Apple logo. Below this is a link to "apple.com" with a globe icon. At the bottom, there is a short paragraph: "Apple Inc. is an American multinational technology company headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, computer software, and online services."

• **Search engine**



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

- **Scraper**

- Data Scraper, 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 results, different types of answers are provided including videos, images, featured 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

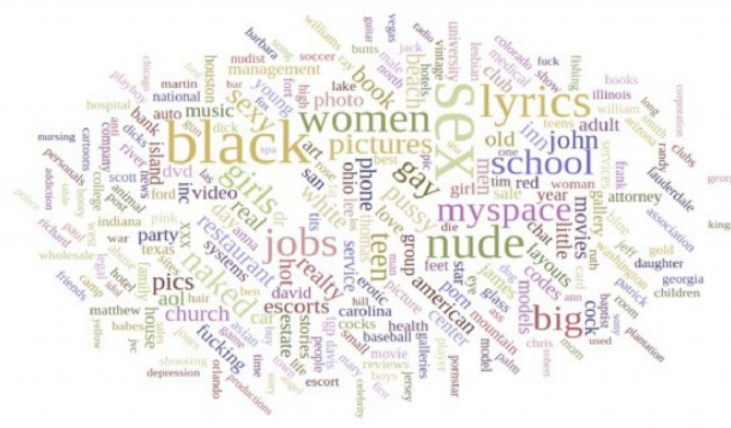


Intent Detection Result

- WordCloud Representation



(a) Cluster0



(b) Cluster1



(c) Cluster2

- **Cluster0:** {New, Best, American} + {Tax, Business, car, Education, College, University}
- 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)			Precision	Recall
		Informational	Local/place	Sexual/Racism		
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

