

Modélisation des Réseaux et Web

Ensemble 4

Ale Abdo (LISIS-UPE)

Un moteur de recherche moderne

1. Avant la requête

2. Réponse à une requête

3a. Résultats

« 10 liens ordonnés »

Un moteur de recherche moderne

1. Avant la requête

Crawler

Analyse de pages

- liens
- contenus

Indexer les contenus

Créer le **réseau** de liens

Calculer les scores

2. Réponse à une requête

(croiser index et requête)

Ramasser les pages candidates

Combiner les scores

Scores

Génériques

- PageRank, score du site

Ciblés à la requête

- fréquence

Ciblés à l'utilisateur

- langue

3a. Résultats

« 10 liens ordonnés »

Un moteur de recherche moderne

1. Avant la requête

Crawler

Analyse de pages

- liens
- contenus
- **sémantique**

Indexer les contenus

- **analyse contextuel**

Créer le réseau de liens

Calculer les scores

Habitudes de l'utilisateur

2. Réponse à une requête

Interprétation

- **entités, synonymes**

Ramasser les pages candidates

Combiner les scores

Scores

Génériques

- PageRank, score du site

Ciblés à la requête

- fréquence, **contexte,**
proximité physique

Ciblés à l'utilisateur

- langue, **intérêts**

3a. Résultats

« 10 liens ordonnés »

Un moteur de recherche moderne

1. Avant la requête

Crawler

Analyse de pages

- liens
- contenus
- sémantique

Indexer les contenus

- analyse contextuel

Créer le réseau de liens

Calculer les scores

Habitudes de l'utilisateur

2. Réponse à une requête

Interprétation

- entités, synonymes

Ramasser les pages candidates

Combiner les scores

Scores

Génériques

- PageRank, score du site

Ciblés à la requête

- fréquence, contexte,
proximité physique

Ciblés à l'utilisateur

- langue, intérêts

3a. Résultats

« 10 liens ordonnés »

Un moteur de recherche moderne

1. Avant la requête

Crawler

Analyse de pages

- liens
- contenus
- sémantique

Indexer les contenus

- analyse contextuel

Créer le réseau de liens

Calculer les scores

Habitudes de l'utilisateur

2. Réponse à une requête

Interprétation

- entités, synonymes

Ramasser les pages candidates

Combiner les scores

Scores

Génériques

- PageRank, score du site

Ciblés à la requête

- fréquence, contexte, proximité physique

Ciblés à l'utilisateur

- langue, intérêts

3a. Résultats

« 10 liens ordonnés »

3b. Retours

Clics

Navigation

Nouvelles requêtes

3c. Testes

Testes A/B (utilisateurs)

Évaluateurs (employés)

Objectifs

Relevance Qualité

Temps

Un moteur de recherche moderne

1. Avant la requête

Crawler

Analyse de pages

- liens
- contenus
- sémantique

Indexer les contenus

- analyse contextuel

Créer le réseau de liens

Calculer les scores

Habitudes de l'utilisateur

2. Réponse à une requête

Interprétation

- entités, synonymes

Ramasser les pages candidates

Combiner les scores

Scores

Génériques

- PageRank, score du site

Ciblés à la requête

- fréquence, contexte, proximité physique

Ciblés à l'utilisateur

- langue, intérêts

3a. Résultats

« 10 liens ordonnés »

3b. Retours

Clics

Navigation

Nouvelles requêtes

3c. Testes

Testes A/B (utilisateurs)

Évaluateurs (employés)

Objectifs

Relevance Qualité

Temps

Systemes de recommandation

- N suggestions à offrir pour une personne
- On connait son profil et comportement antérieur
- On connait le profil et comportement de plusieurs autres
- Quels options l'offrir et dans quel ordre ?

Systemes de recommandation

- Similarités
 - Entre personnes, entre objets
 - Sources intérieures, extérieures
- Score
 - Contribution personnalisé (combine les sources de similarité)
 - Contribution généraliste (autres informations sur la personne ou objet)
- Retours
 - Clics, achats, commentaires, évaluations
- Apprentissage automatique
- Testes A/B, évaluateurs professionnels

Systemes de recommandation

Systemes de recommandation

À vos réseaux

Web, données et interfaces

- Modulaire :
 - contenu (HTML5)
 - style (CSS3)
 - Interactivité (ECMAScript2019)
 - structure (standards spécifiques)
- Sémantique :
 - construction et usage d'ontologies (RDF, Schema.org, Wikidata)
 - inférence de structures et contenus (« web mining »)

Web, données et interfaces

- « Open » : respect à la modularité avec des standards et formats publiquement documentés
- « Linked » : existence et usage de ontologies qui permettent l'interopérabilité entre jeux de données
- « Linked Open » : Linked + Open + ontologies publiquement documentés

Web, données et interfaces

- Architectures intentionnelles optimisées
 - comportement, testes A/B
- Équilibre entre raffinement et pouvoir statistique

Réseaux et classification

Classification, communautés, clusters, hiérarchies

Réseaux et classification

Modularité et motifs assortatifs

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

Réseaux et classification

Modèles à blocs stochastiques