

Résumé Détaillé du Cours : Introduction à la Recherche d'Information (RI)

1. Objectifs du Module

- Présenter les notions de base de la recherche d'information.
- Apprendre les différents modèles de recherche d'information.
- Se familiariser avec les techniques de base de la RI.
- Mettre l'accent sur les méthodes d'indexation, de traitement des requêtes et les métriques d'évaluation des performances.

2. Introduction à la Recherche d'Information

Définition de la RI :

- La RI est une branche de l'informatique qui s'intéresse à l'acquisition, l'organisation, le stockage, la recherche et la sélection d'information.
- Elle utilise des méthodes et techniques pour fournir des informations pertinentes à un utilisateur.

3. Historique de la RI

Années 1940 : Apparition des premiers systèmes de RI, principalement utilisés dans les bibliothèques.

Années 1950 : Introduction du modèle booléen et premières expérimentations sur des petites collections de documents.

Années 1960-1970 : Développement du système SMART par G. Salton, méthodologies d'évaluation et création de corpus de test.

Années 1980 : Intégration de l'intelligence artificielle dans la RI (systèmes experts).

Années 1990-1995 : Impact d'Internet, élargissement des problématiques de RI (documents multimédias).

Années 2000 :

- Analyse de liens (Google).
- Extraction d'information, réponses à des questions, indexation multimédia (images, vidéos, audio).
- Recherche multilingue, résumé automatique de documents.

4. Objectifs de la RI

- Identifier et exploiter l'information contenue dans des documents (texte, son, image) en réponse à une requête utilisateur.
- Trouver des documents pertinents dans des ensembles volumineux (Big Data).
- Minimiser les documents non pertinents retournés par le système.
- Gérer des contenus non structurés ou semi-structurés.

5. Domaines d'Application de la RI

- Internet : Moteurs de recherche (Google, Yahoo, Bing), forums, blogs, actualités.
- Entreprises : Recherche interne (entreprise search).
- Bibliothèques numériques : Accès à des documents numérisés.
- Domaines spécialisés : Médecine, droit, littérature, chimie, mathématiques, brevets, logiciels.
- Recherche personnelle : Recherche sur les ordinateurs personnels (Yahoo! Desktop Search).

6. Acteurs de la RI

- La collection de documents : Ensemble des informations exploitables.
- L'utilisateur : Celui qui exprime un besoin d'information et bénéficie des résultats.
- Le système de RI : Outil qui retrouve les documents pertinents pour l'utilisateur.

7. Système de Recherche d'Information (SRI)

Définition : Programme informatique qui sélectionne des informations pertinentes en réponse aux besoins des utilisateurs.

Concepts de base :

Donnée vs Information vs Connaissance :

- Donnée : Chaîne de caractères avec des valeurs associées.
- Information : Signification des données, compréhensible par l'utilisateur.
- Connaissance : Information découverte, comprise et partagée par une communauté.
- Formes de l'information : Texte, images, sons, vidéos, graphiques.
- Propriétés de l'information : Structure (structurée, semi-structurée, non structurée), hétérogénéité (langage, média, structures).

8. Notions Clés dans un SRI

- Document : Toute unité pouvant constituer une réponse à une requête (texte, page web, image, vidéo, etc.).
- Corpus : Recueil de documents (textes, sons, images) utilisé pour des analyses scientifiques.
- Requête : Expression du besoin d'information de l'utilisateur.
- Pertinence :
 - Pertinence utilisateur : Satisfaction de l'utilisateur par rapport aux documents retournés (subjective).
 - Pertinence système : Mesure algorithmique de la similitude entre un document et une requête (objective).

Objectif du SRI : Rapprocher la pertinence système de la pertinence utilisateur.

9. Processus de SRI

- Requête : Expression du besoin d'information.
- Traitement : Indexation et recherche (modèles vectoriels, probabilistes).
- Indexation : Organisation des documents pour faciliter la recherche.
- Résultat : Retour des documents pertinents à l'utilisateur.

10. Références Bibliographiques

[1] <http://b3d.bdpedia.fr/ri-intro.html>

[2] https://www.prologism.fr/wp-content/uploads/2020/08/PROLOGISM_BigData-GestionDonneesSemi-Structures-SOL.pdf

[3] https://staff.univ-batna2.dz/sites/default/files/djemane-azeddine/files/semistrucdata_chap1.pdf

Structure Organisée avec Couleurs

Objectifs du Module : 

Introduction à la RI : 

Historique de la RI : 

Objectifs de la RI : 

Domaines d'Application : 

Acteurs de la RI : 

Système de RI (SRI) : 

Notions Clés dans un SRI : 

Processus de SRI : 

Références : 

Ce résumé détaillé et structuré permet de comprendre les concepts clés de la recherche d'information, son historique, ses objectifs, et les acteurs impliqués dans le processus.