

Université Jean Monnet, Saint-Étienne
Faculté des Sciences et Techniques de Saint-Étienne
Laboratoire Hubert Curien, UMR CNRS 5516

Recherche d'Information (RI)

Mathias Géry

Mathias.Gery@univ-st-etienne.fr



Laboratoire Hubert Curien

- Département « Optics, Photonics and Surfaces ».
- Département « IA, Sécurité, Image », 3 équipes :
 - « Image Science & Computer Vision ».
 - « Sécurité des Systèmes Embarqués et Architectures Matérielles ».
 - « **Machine Learning** » (**Rémi Emonet**).
 - Projet « Foundations & Theory of Machine Learning » (E. Morvant).
 - Projet « Responsible & Thrustworthy Machine Learning » (C. Langeron)



Information Retrieval (IR)

Information overload (surcharge informationnelle) (infobésité)

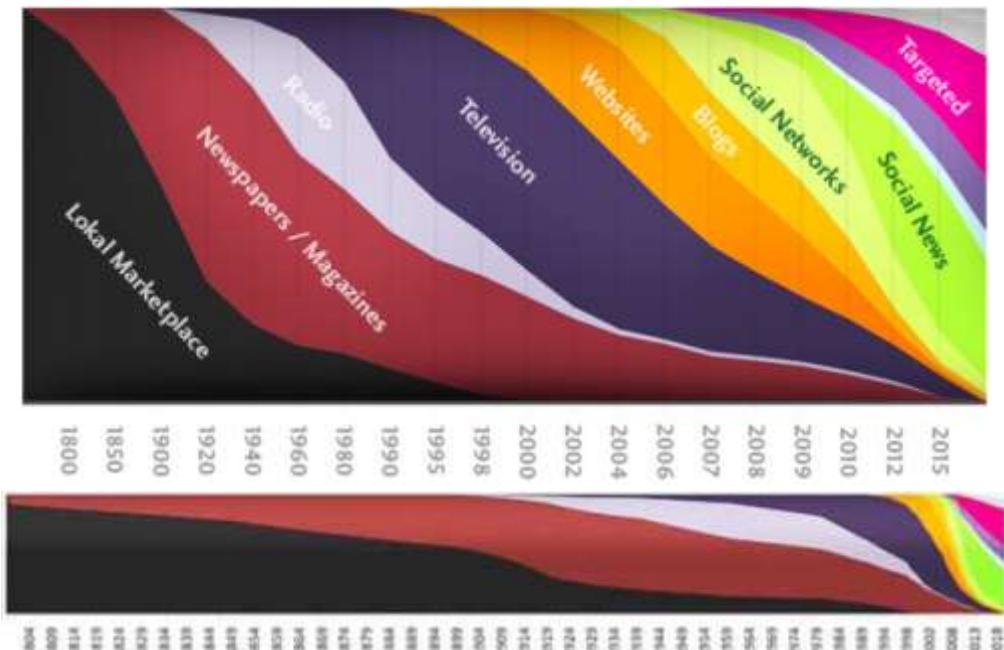
THE INTERNET IN **2023** EVERY MINUTE



Source



Gradual changes of channels



Petr Sojka, Hinrich Schütze



Information Retrieval (IR)

- Information Retrieval (IR) is **finding material** (usually documents) of an **unstructured** nature (usually text) that satisfies an **information need** from within **large collections** (usually stored on computers).
- These days we frequently think first of **web search**, but there are many other cases:
 - E-mail search,
 - Searching your laptop,
 - Corporate knowledge bases,
 - Legal information retrieval,
 - Etc.



“Information Retrieval and Web Search”, Stanford

Recherche d'Information (RI) : organisation du cours

- Mathias Géry, Mathias.Gery@univ-st-etienne.fr
- Au programme : 13 séances x 3 heures.
 - Cours (anglais / français).
 - TD : exercices, articles de recherche.
 - TP / Projet (en équipe).
- Planning :
 - Jeudi 14h (en général).
 - Attention aux modifications de dernière minute !
→ vérifiez ADE !
- ENT / Moodle :
 - « [M2 DSC Recherche d'Information 2025-26](#) »
 - Diapos, exercices, ressources pour le projet.
 - Rendu de TD / TP / Projet.
 - 19 inscrits au 17/09/2025 (sur 22 ?).

Recherche d'Information (RI) : évaluation

- Exam (50%) : 12 février (dernière des 13 séances).
 - Projet (50%) :
 - « Moteur de recherche de documents structurés.
 - Participation à une mini-compétition de RI. »
-
- À réaliser (en équipe de 3) essentiellement entre les séances,
 - $22 = 7$ équipes + 1 redoublant.
 - Pendant les séances : discussions, « TP » (PC portable ?),
 - Étapes intermédiaires (chaque séance),
 - Rendu final : 13 février (« hard deadline »).
 - ➔ note « pratique » de l'UE « Recherche d'Information ».

Recherche d'Information (RI) : ressources

- Courses:
 - “[Information Retrieval and Web Search](#)”, Stanford.
 - “[Intelligent Information Retrieval](#)”, DePaul.
- Books:
 - [Introduction to Information Retrieval](#), by C. Manning, P. Raghavan, and H. Schütze ([pdf](#)).
 - Recherche d'information, Applications, modèles et algorithmes, by M-R. Amini and E. Gaussier, 2013.
 - [Modern Information Retrieval](#), the concepts and technology behind search, by R. Baeza-Yates and B. Ribeiro-Neto.
 - [Managing Gigabytes](#), by I. Witten, A. Moffat, and T. Bell.
 - [Search Engines: Information Retrieval in Practice](#), by B. Croft, D. Metzler, and T. Strohman ([pdf](#)).
- IR resources : [Glasgow](#) (collections, systems, books, etc.)
- Software:
 - IR Systems: [Terrier](#), [Apache Lucene](#) ([solr](#)), [Elasticsearch](#), [Lemur](#).
 - Others: [Apache Nutch](#), [NLTK](#), [scikit-learn](#), etc.



Information Retrieval: Outline

- 0) Introduction
- 1) IR Model & Boolean Retrieval
- 2) Pre-Processing & Dictionary
- 3) Ranked Retrieval
- 4) Experimental Evaluation of IR
- 5) Structured Retrieval
- 6) Link Analysis for IR