acronym

Mike Grehan | CMO & Managing Director

SEO pioneer, author, world-traveler and keynote speaker, Champagne connoisseur and consummate drinking partner to the global digital marketing community. Former publisher of SearchEngineWatch.com and ClickZ.com, and producer of the industry's largest search and social marketing event, SES Conference & Expo. Proud to be Chairman of SEMPO the largest global trade Association for search marketers. And equally proud to be CMO & Managing Director of the darned finest search marketing agency in the known universe. Fun fact: Used to be a popular radio DJ in the UK… Before finding a proper job.

Chairman

SEMPO

The Largest Nonprofit Trade Organization in the World
Serving the Search and Digital Marketing Industry.

acm

The World's Largest Computing Society

Senior Member

Global Marketing Day
October 30th 2019
New York City
SEMRUSH
Wall Street Station
Uptown & The Bronx
4
5
Underpass to Downtown & Brooklyn

acronym

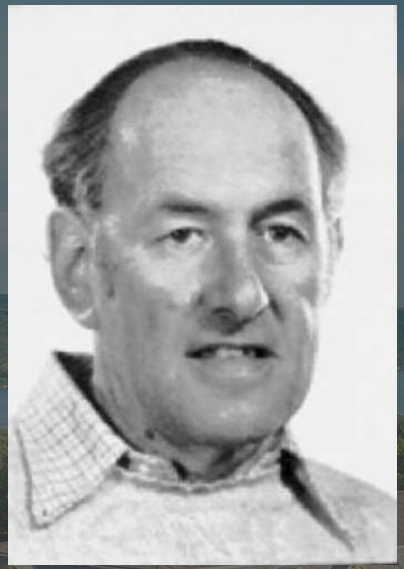
It's All About The Data, Baby!

What's the difference between information retrieval and data retrieval?

261 MW
97 FM
IN TUNE
WITH THE NORTH EAST
METRO
RADIO
Mike Grehan

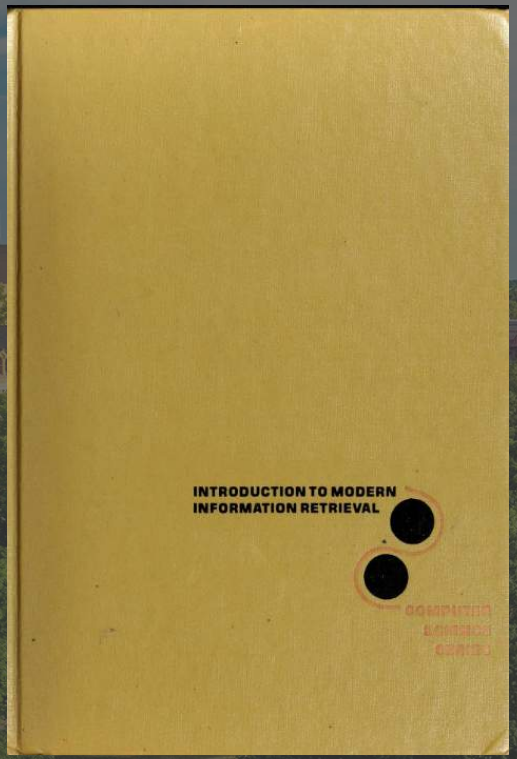
1983

Gerard Salton | Father of Modern Information Retrieval



1983

Cornell University



What is "information retrieval?"

"Information retrieval is a field concerned with the structure, analysis, organization, storage, ***searching and retrieval of information***."

Gerard Salton - Father of Modern Information Retrieval

acronym

What is “data retrieval?”

“A database management system is software that enables the creation, maintenance and use of large amounts of data. ***Databases contain highly structured data*** subdivided into fields and stored in a collection of files containing records on a common theme which is queried by expert users (programmers) using formal query languages."

Mike Grehan – Father of five, grandfather to six.

acronym

Comparison of Database Systems and Information Retrieval

	Database Systems	Information Retrieval
Data type	Numbers, short strings	Text
Foundation	Algebraic/logic based	Probabilistic/statistics based
Search paradigm	Boolean retrieval	Ranked retrieval
Queries	Structured query languages	Free text queries
Evaluation criteria	Efficiency	Effectiveness (user satisfaction)
User	Programmer	Nontechnical person

Index Corpus Taxonomy

The web is an ocean of ***human generated unstructured data*** permanently shifting With tides of change. It's a major challenge for Google to gather, process and Index unstructured data from the gazillions of websites vying for visibility on the World Wide Web.

acronym

Information Retrieval and Web Search

Three main categories:

Page Content

Automatically **crawl, classify and cluster** web pages according to topics.

Hyperlink Analysis

Crucial to discovering **important web pages.** Links also help to **identify communities** that share common interests.

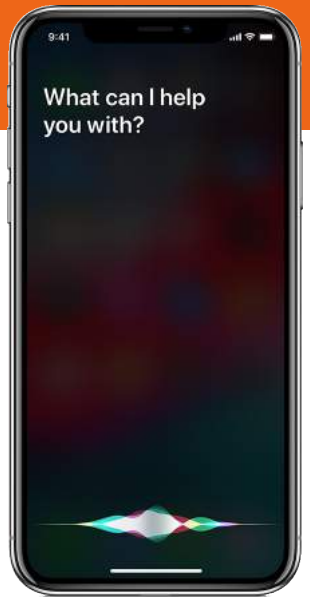
Usage Data

Usage mining refers to access patterns and click curves often referred to as **end user data**.

E A T

Natural Language Processing (NLP)

Most research in natural language processing (or computational linguistics) aims to capture the meaning of text. It is meant to yield deeper representations that are closer to meaning and may be exploited in real-world applications.



End-user applications include:

- Machine translation
- Text summarization
- Sentiment analysis
- Voice activated queries
- Question answering

Information extraction (IE):

The automatic extraction of structured information such as entities, relationships between entities, and attributes describing entities from unstructured sources.

BERT

"I'm already over it....
And you should be too!"

acronym

The Semantic Web

The Semantic Web (coined by World Wide Web inventor Sir Tim Berners-Lee) is relatively new compared to Information Retrieval, Databases and Natural Language Processing. The original Web is a medium of documents for people, but the Semantic Web is meant to be a Web of "actionable information" as in an environment that enables intelligent agents to carry out sophisticated tasks for users.

Sir Tim Berners Lee

A language that expresses both data and rules for reasoning about the data.

A Web of relations between resources denoting real world objects:

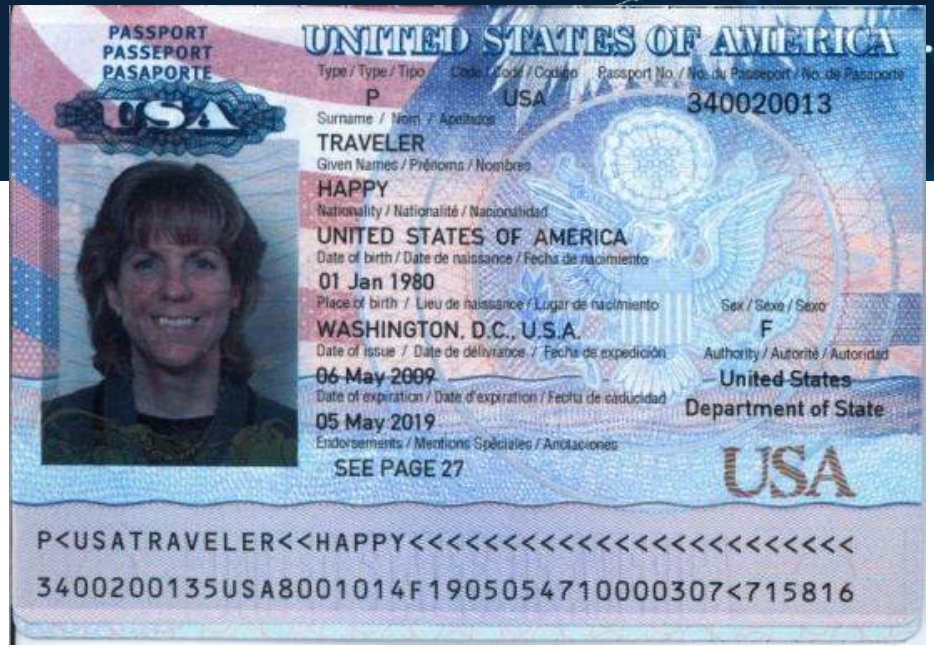
- People
- Places
- Events

Knowledge representation:

- RDF/JSON
- RDF Schema
- Simple Knowledge Organization System (SKOS)
- Web Ontology Language (OWL)

acronym

Describe Machine Readable



acronym

What is “entity search?”

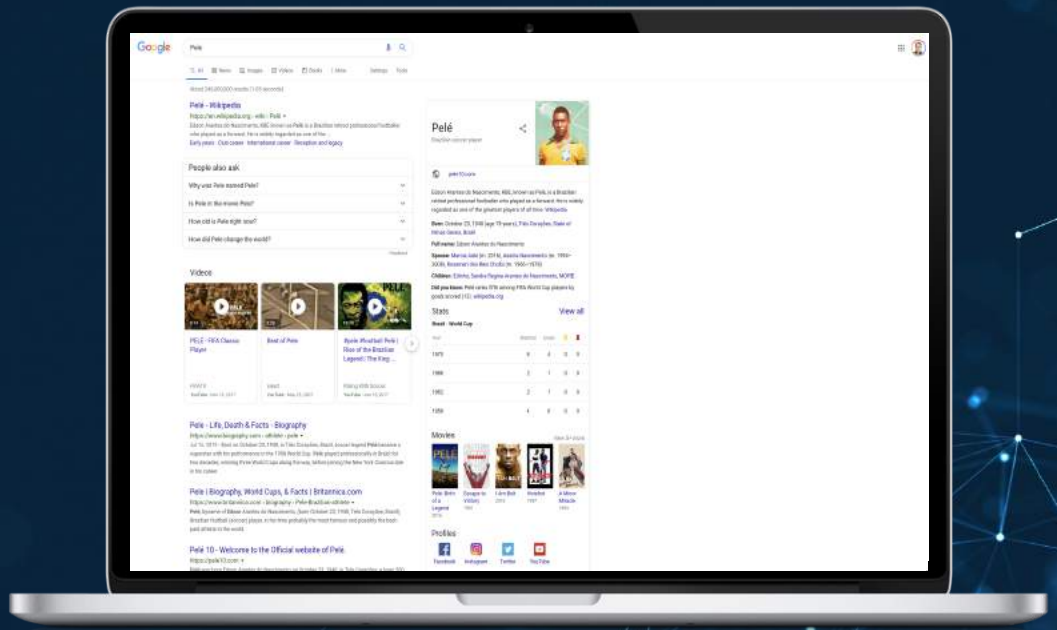
Informally, an entity is a “thing” or “object” that can be referred to. Common types of entities include, e.g., people, organizations, products, locations, and events.

Author - **Krisztian Balog**

**An entity is an object or concept in the real world that can be distinctly identified.*

acronym

Understand how structured data works.

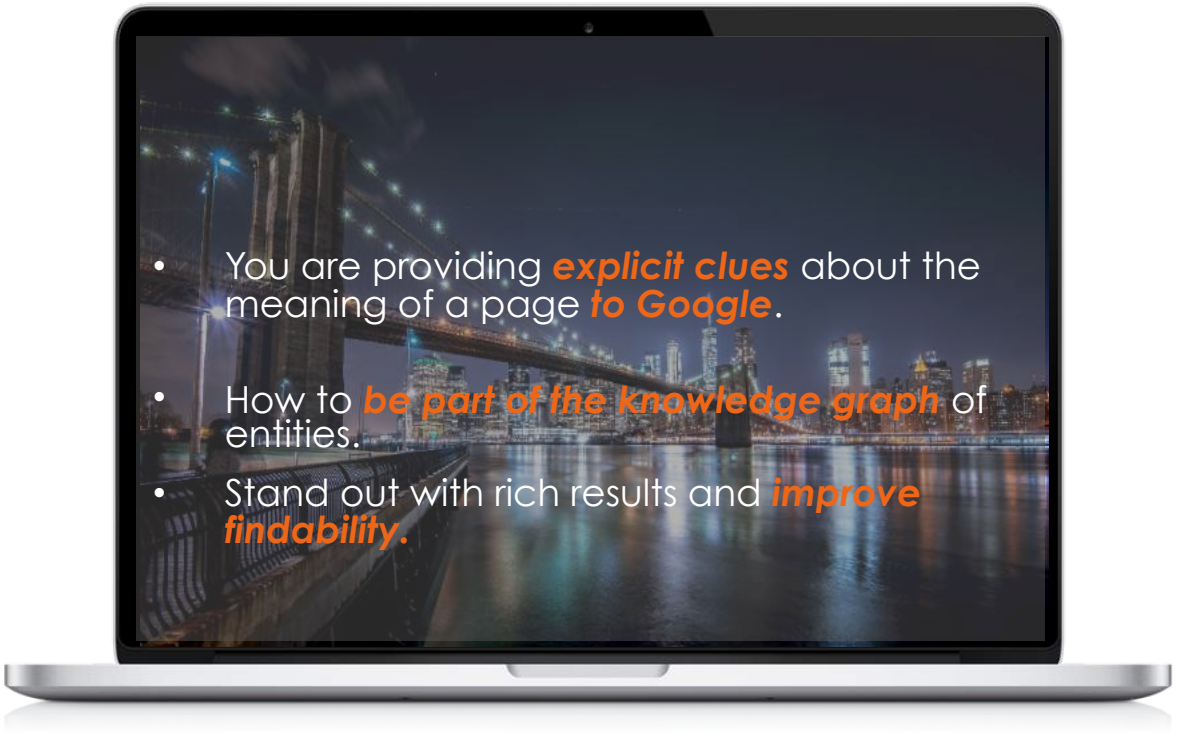


Google Search works hard to understand the content of a page. You can help us by providing explicit clues about the meaning of a page to Google by including structured data on the page.

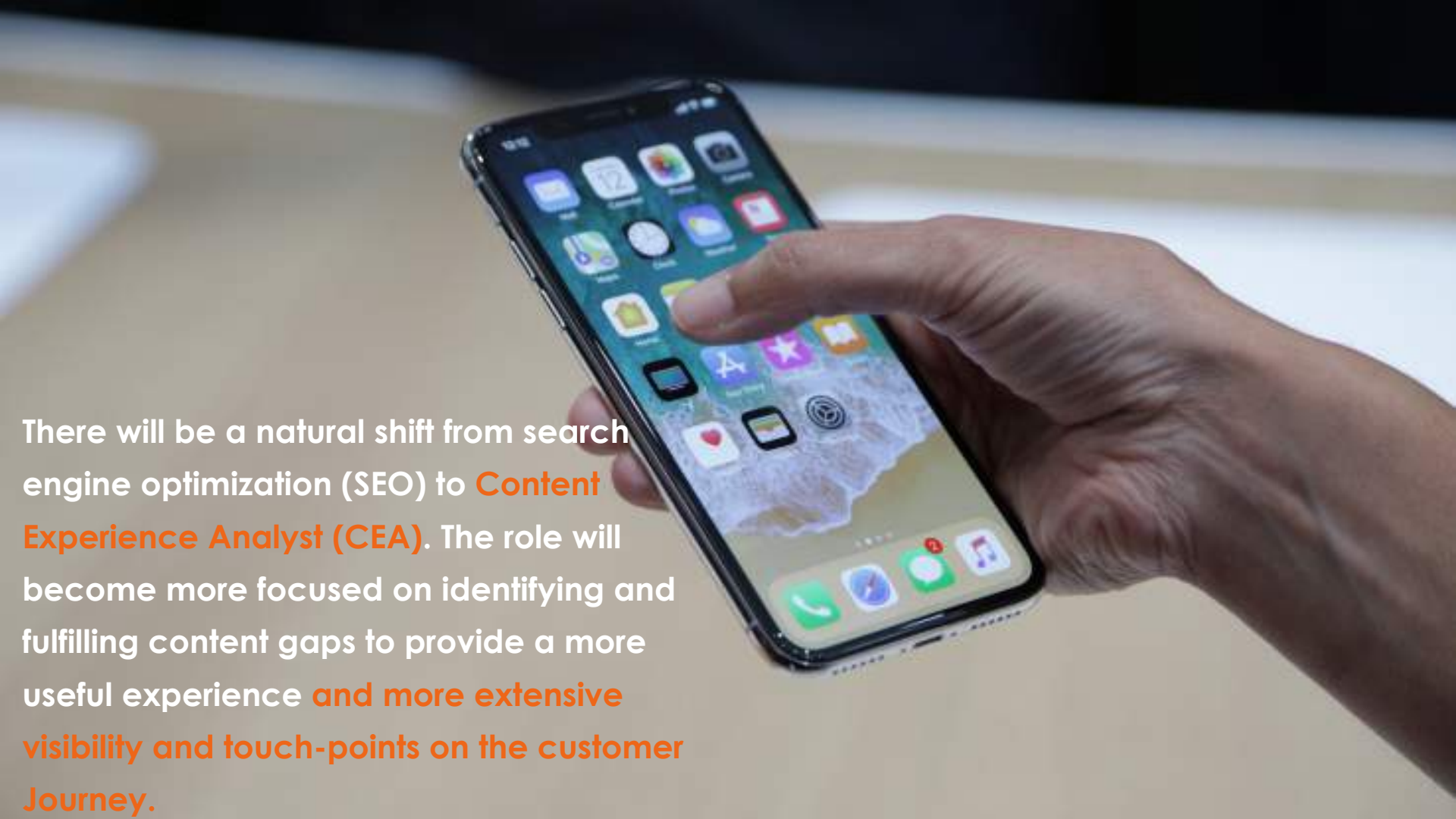
Structured data is a standardized format for providing information about a page and classifying the page content; for example, on a recipe page, what are the ingredients, the cooking time and temperature, the calories, and so on.

What Have We Learned (1)?

What Have We Learned (2)?



What Have We Learned (3)?



There will be a natural shift from search engine optimization (SEO) to Content Experience Analyst (CEA). The role will become more focused on identifying and fulfilling content gaps to provide a more useful experience and more extensive visibility and touch-points on the customer Journey.

Mike Grehan

Data is not information,

Information is not knowledge,

Knowledge is not understanding,

Understanding is not wisdom.

Author, Clifford Stoll

THANK YOU