# Elasticsearch Training Session 1

Rajan Manickavasagam

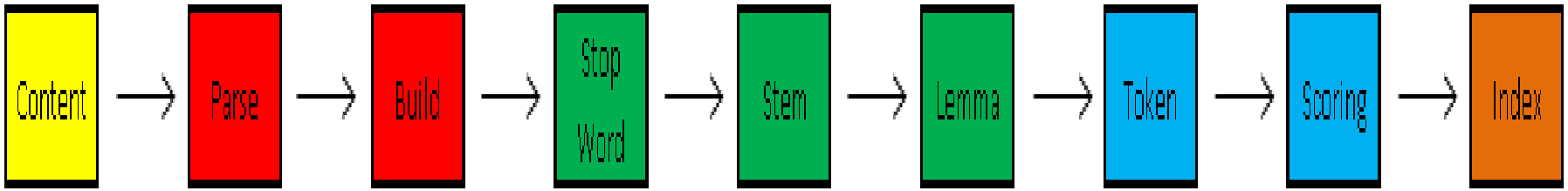
# Agenda

- History
- Overview
- Types of Search
- Key Concepts
- Introduction to Elasticsearch

# History

- 1950's – Douglas Engelbart created the AHI (Augmented Human Intellect) program at Stanford Research Institute. He is more famous as the inventor of the mouse.
- 1960's – IBM created a full text search system called STAIRS (Storage and Information Retrieval System). This was the first enterprise search product.
- 1980's – Verity, spun off from Advanced Decision Systems came up with real time indexing. It eventually morphed into Inktomi and created the UltraSeek product. In 2005, Verity was acquired by Autonomy and Autonomy in turn by HP.
- 1991 – The first web page was online from CERN.
- 1994 to 1997 – Various web search engines like Altavista, Excite, Lycos, Yahoo rose to prominence.
- 1997 – Page Rank & Hypertext Induced Topic Search (HITS) algorithms created by Google.

# Overview

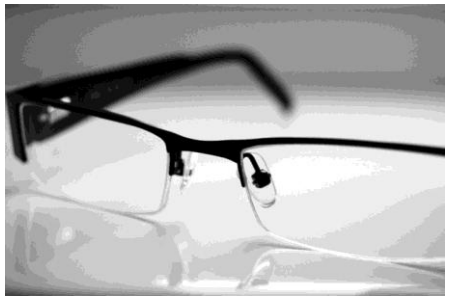

Content
Parse
Build
Stop Word
Stem
Lemma
Token
Scoring
Index

# Types of Search

Text

Voice

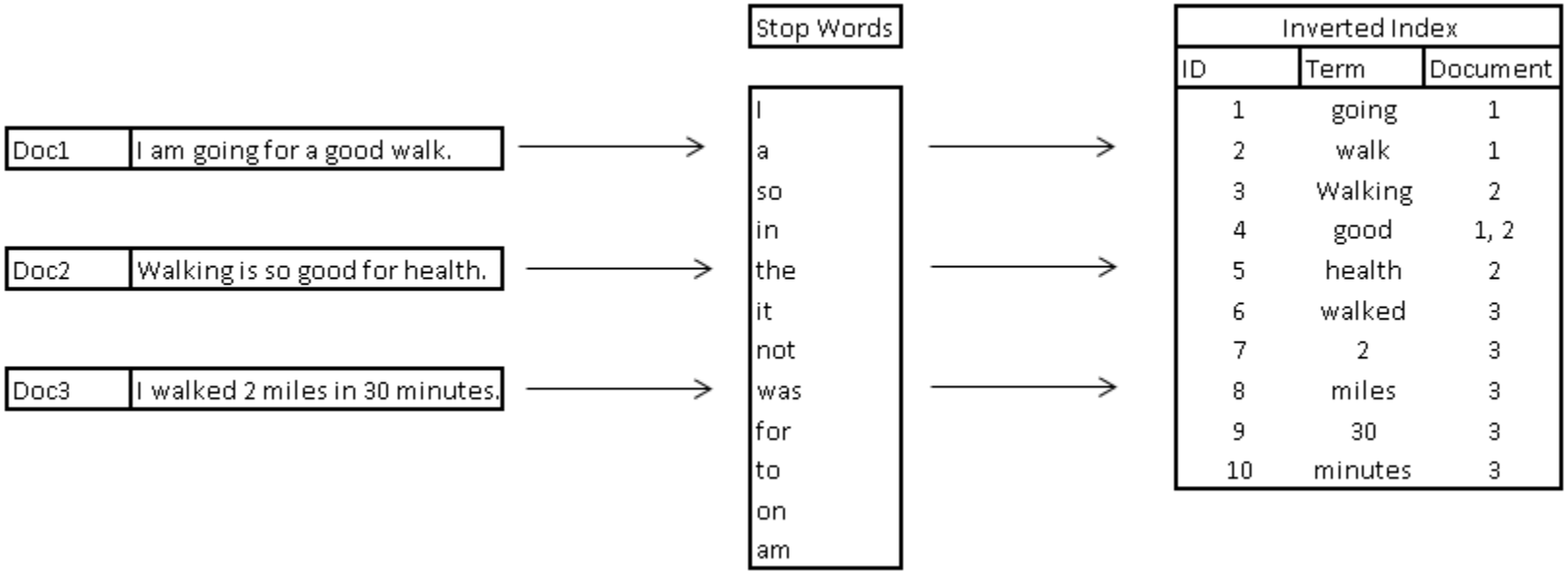
Visual



Semantic

# Key Concepts

- Inverted Index



# Introduction to Elasticsearch

- Created by Shay Bannon in 2010.
- It is an open source project licensed under Apache 2 license.
- It is a RESTful and distributed search engine.
- Index – Data is stored in indices.
- Document - One ore more documents make up an index. Each document is a JSON.
- Type – Each document is associated to a type.
- Mapping – Defines the indexing strategy for documents.
- Node – Each instance is called a node.
- Cluster – Multiple nodes make up a cluster.
- Shard – Data in an index can be partitioned across shards. This improves indexing. All create, update & delete go to primary shard.
- Replicas – Replicas of index are provided for better read performance and availability. All reads and queries can go to primary/replica.
- Gateway – Cluster information is persisted in the gateway.

# Introduction to Elasticsearch

- Default values available to start the cluster.
- It works in distributed mode by default.
- It has a peer-to-peer architecture.
- Can scale up and scale out.
- Indexing and searching happens in near real time.
- All operations are available through REST API's.
- All indices/shards are internally, independent Lucene instances.

# References

- Guide - https://www.elastic.co/guide/en/elasticsearch/guide/current/index.html
- Reference - https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html
- Plugins - https://www.elastic.co/guide/en/elasticsearch/reference/current/modules-plugins.html
- Java API - https://www.elastic.co/guide/en/elasticsearch/client/java-api/current/index.html
- .Net API - https://www.elastic.co/guide/en/elasticsearch/client/net-api/current/index.html