Search Terminology

in SharePoint and Office 365

Agnes Molnar, Search Explained

2016

Search Terminology in SharePoint and Office 365

© 2016 All rights reserved.

Published by Search Explained

http://SearchExplained.com

E-book edition created 2016.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means – for example, electronic, photocopy, recording – without the prior written permission of the publisher. The only exception is brief quotations in reviews, with a link to the book and/or http://SearchExplained.com.

All trademarks, service marks, trade names, product names and logos appearing in this book are the property of their respective owners.

This e-book is an independent publication and is not affiliated with, nor has it been authorized or sponsored, by Microsoft Corporation.

Introduction

SEARCH EXPLAINED
SEARCH. FIND. BE SUCCESSFUL.

The terminology of Enterprise Search is a special one. We talk about it, discuss it, even have debates about it - but the exact phrases and expressions might have "blur" understandings.

Based on my experience, Enterprise Search discussions almost always have to be started with carifying the terminology, which may take valuable time.

The main goal of this mini e-book is to help everyone working with Search to have a common understanding, and that the discussions to be started with the real, valuable conversations.

To your Search success,

Agnes Molnar

Search Explained, 2016

http://SearchExplained.com

Agnes.Molnar@SearchExplained.com

Search Scenarios

Search. Discovery.

Search

Search is an action with the intent of getting relevant and useful content which helps to get done a particular task or job.

Search is usually initiated by some user interaction: in most cases, entering a query, choosing some filtering options, or navigating by a menu.

Search can be intended for a single document or small set of documents, or it can be exploratory search when the user wants to "explore" the available content and collect the useful pieces during his/her "journey".

Discovery

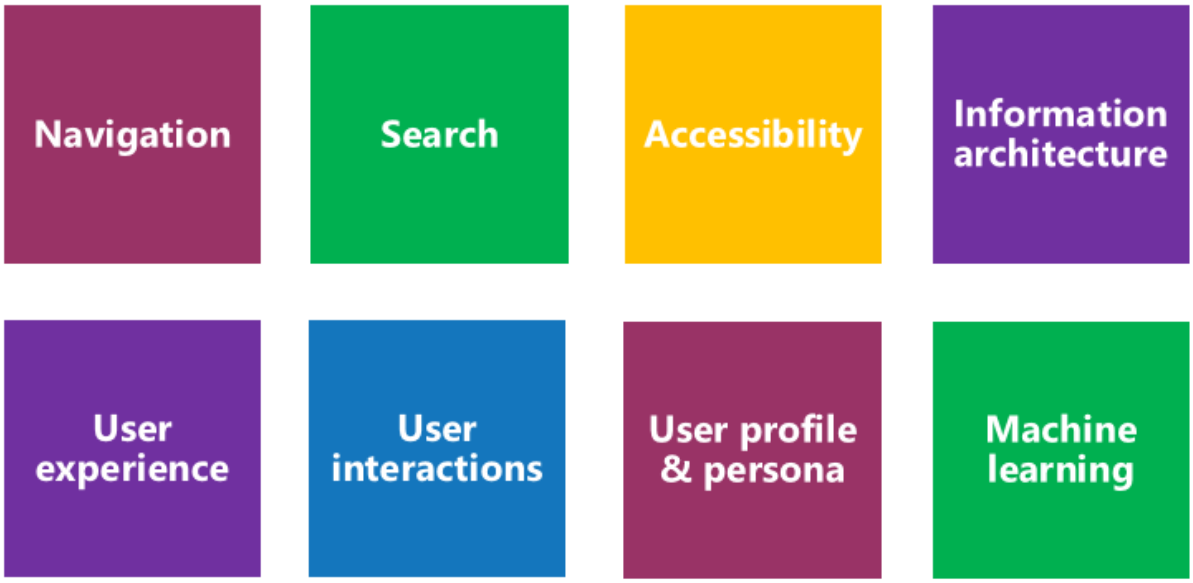
By leveraging statistical, data mining and machine learning techniques, discovery tools provide personalized content suggestions to every user. Content discovery is an important part of findability because users can find (discover) content that they do not even know about this way, without proactively searching for it.

For example, Delve, powered by Office Graph, is a discovery tool in Office 365.

Search Scenarios

Findability.

Findability



Findability is the capability of finding the proper document(s) or other content piece(s) according to the context and the user's current intent.

Findability includes *Search*, but it is much more than that. It also consists of navigation, information architecture, user experience, and accessibility as well as *discovery* tools and machine learning techniques.

Search Scenarios

Web Search. Enterprise Search. Real-Time Search. Semantic Search.

Web Search

Search action, initiated on the internet, with the intent to find (public) web content.

Enterprise Search

Enterprise Search connects people to the information they need to get their *jobs* done. The goal is to make enterprise content findable, both inside and outside the organization.

Enterprise Search is controlled, maintained, and owned by the organization. Therefore, customizations and custom developments are possible, and in most cases also needed.

Real-Time Search

Real-time search provides real-time data, rather than indexed content. Real-time search applications are used in big data scenarios, on social networks, etc. They are not typically used in enterprise search scenarios, but might be required. Due to the crawling model, it is not feasible in SharePoint / Office 365.

Semantic Search

Semantic Search's goal is to improve **search** relevancy by understanding the searcher's intent, and the context of searching. It uses natural language techniques, concept matching, synonyms, query generalization, and specialization, etc. It is widely used for discovery scenarios.

Search Scenarios

Data. Information. Knowledge.
Information Architecture. Knowledge Management.

Data

Facts and figures of the world. Data is stored, but not organized.

Information

Contextualized, categorized and organized data. Information is sometimes calculated based on one or more data values. Information always has a *purpose*.

Knowledge

Understanding, experience, insight, intuition and information put into *context*.

Information Architecture

The art and science of organizing and labeling the content (documents, websites, blog posts, database entries, etc.) in order to support findability and usability.
Information Architecture has a significant effect on Search and Findability.

Knowledge Management

The process of capturing, developing, sharing and effectively using organizational knowledge. It consists of processes, technology, culture, structure, community considerations.

Topology

On-premises Search. Cloud Search. Hybrid Search.

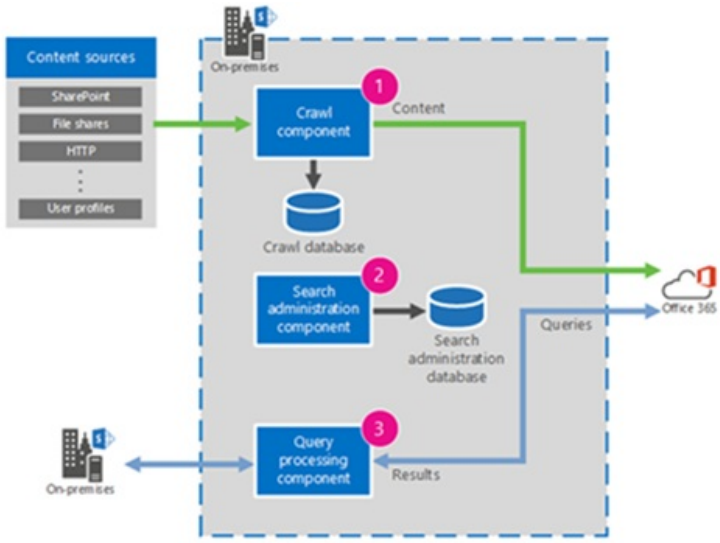
On-premises Search

When a Search Engine is installed on self-hosted servers and virtual machines on-premises, and the content is also on the organization's on-prem infrastructure, we talk about on-premises Search.

Cloud Search

When the content and the Search engine are both in the cloud, for example in Office 365, it is called Cloud Search.

Hybrid Search



Hybrid Search is the combination of on-premises and cloud searches. In most cases, content is located in both on-premises and the cloud, as well as the Search Engine.

Topology

Search Service Application. Cloud SSA.

Search Service Application (SSA)

In SharePoint, Service Applications are responsible for the deployment and configuration of services. This service infrastructure gives us the control over which services are deployed and how services are shared.

Search Service Application is responsible for the configuration of Search, including the Search Infrastructure, Content Sources, Crawl schedule, farm-level Search Schema (metadata), etc.

Cloud Search Service Application (Cloud SSA)

In SharePoint 2016, and in SharePoint with the 2015 August Cumulative Updates, we can configure Hybrid Search to connect on-premises Content Sources to Office 365. This configuration can be done by creating and configuring a Cloud Search Service Application.

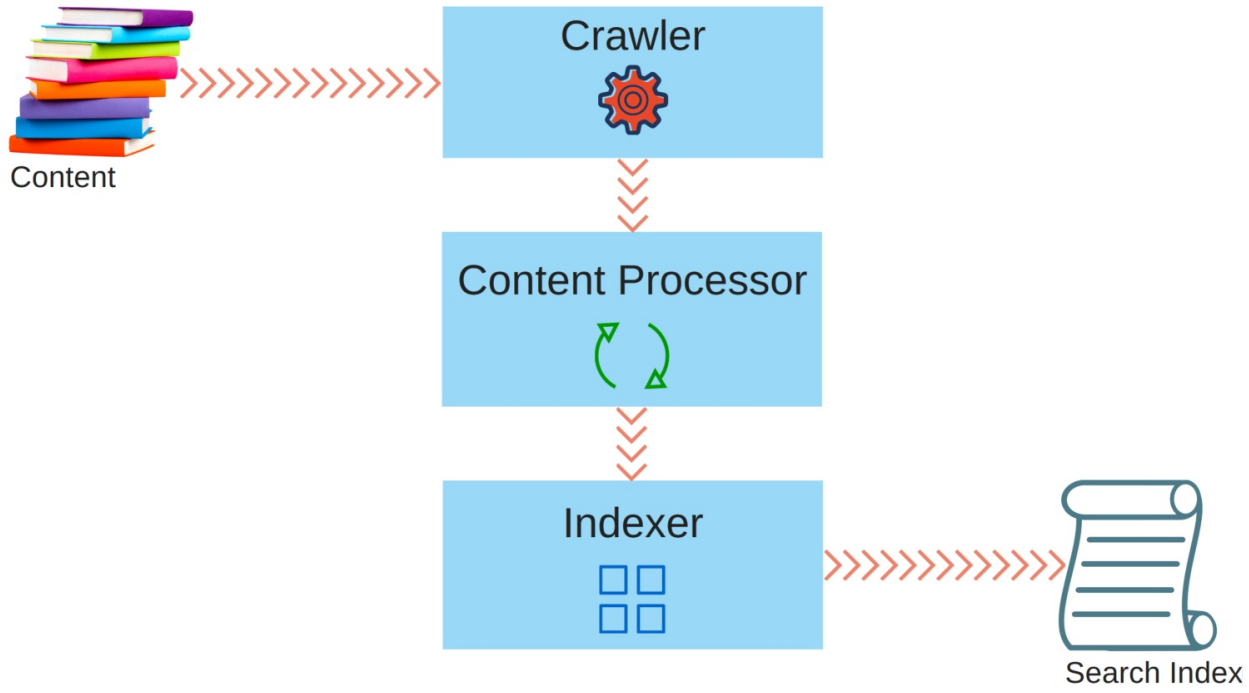
Architecture and Components

Crawling.

Crawling

Crawling is the process of gathering the content to be indexed from the content sources. To retrieve information, the crawl component connects to the content sources by invoking the appropriate indexing connector or protocol handler. After retrieving the content, the crawl component passes crawled items to the Content Processing Component.

The Crawling process, including content sources and crawl schedule, can be configured in the Search Service Application.



Architecture and Components

Full Crawl. Incremental Crawl. Continuous Crawl.

Full Crawl

During Full Crawl, every item in the Content Source gets crawled, regardless of its previous crawl status.

Incremental Crawl

In most cases, we do not need to (re-)index everything in the content source. Incremental Crawl enables the Crawler to crawl only the items which have been newly created or modified since the last crawl (content updates, metadata updates, certain permission updates, deletes, etc.). This way, the crawling and indexing processes take much less time as the unchanged items are not processed.

Continuous Crawl

Continuous Crawl is a dynamic way of crawling SharePoint and Office 365 content. When it is enabled on the Content Source, the Crawler checks the SharePoint changelog regularly (every 15 min, by default) which contains the list of recent changes. If there's any item on the change log, the Crawler takes immediate action on it and sends to the Content Processor.

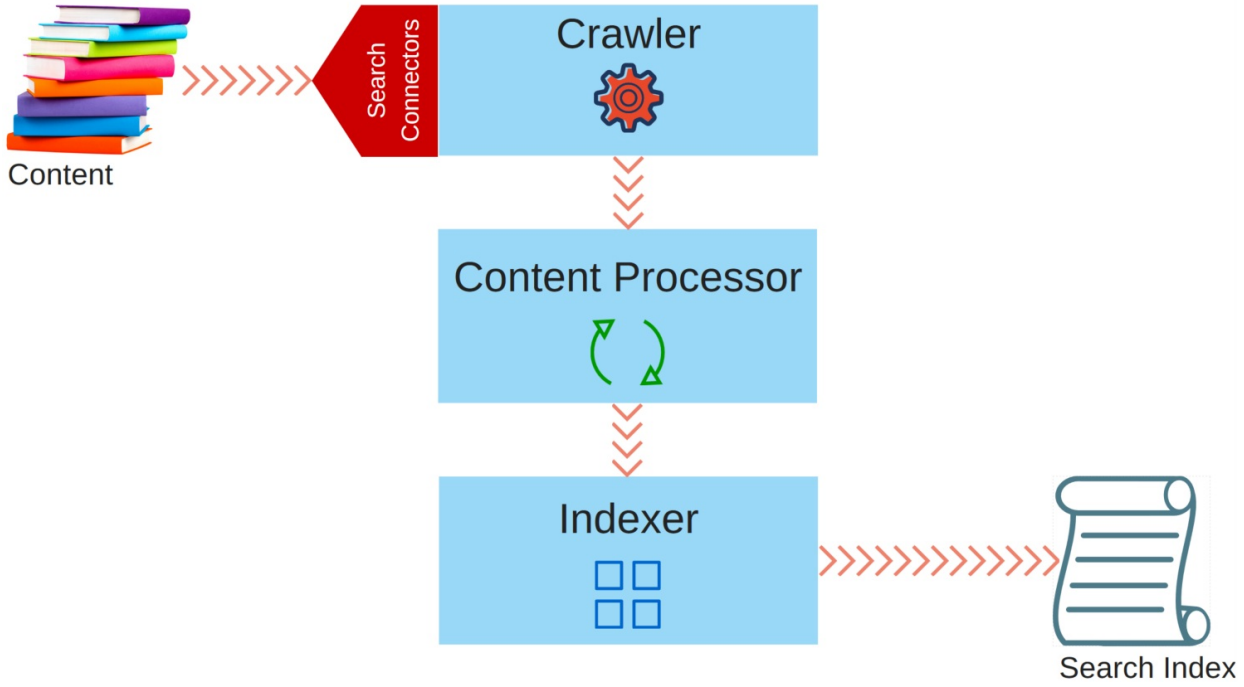
Please note, that Continuous Crawl is not real-time, but it checks the SharePoint change logs often enough to be quick and have good content freshness.

Architecture and Components

Search Connectors.

Search Connectors

Search Connectors enable the Crawler to connect to the source systems' APIs in order to extract the content from for further processing and indexing. SharePoint provides a set of out-of-the-box Connectors, as well as an API on which custom connectors can be built on.



Architecture and Components

Content Processing. Indexing.

Content Processing

The Content Processing Component processes items that have been crawled by the Crawler and then sends these items to the Indexer. It performs operations such as document parsing and property mapping. It also performs linguistics processing and transforms crawled items into artifacts that can be included in the Search Index.

Indexing, Search Index

The processed content goes to the Indexer components which is responsible for storing it in the actual Search Index. Search Index is divided into *index partitions*. Each partition is stored as a set of files. To scale out the search index, new index partitions can be added.

For fault tolerance, *index replicas* can be created of each index partition. Each index replica contains the same information.

Architecture and Components

Search Federation.

Search Federation

There are several cases when we cannot, or don't want to index the source system. Search Federation lets us use a remote Search Index in these cases, instead of crawling and indexing the content into the local Search Index. The remote index has to be compatible with OpenSearch 1.0/1.1 standard.

The most significant benefits of using Search Federation are:

- No need to use local resources for the crawling, content processing and indexing processes in SharePoint Search.
- No need to use resources to store the indexed content in SharePoint index.

At the same time, there are some limitations and disadvantages, too, which have to be evaluated before deciding to use federation:

- No interleaved results (results from the local index and federated locations are separated into separate Result Blocks).
- No refiners are available for the federated results.
- Since the queries are sent to both the local and federated (remote) search indices, the query response time is always longer than by using a single index. The exact additional delay depends on the configuration, network bandwidth, and the remote index's performance.

To work around these limitations, third-party federating connector is needed to be deployed.

It is highly suggested to use crawling connectors whenever possible, otherwise do a Proof of Concept (POC) before make a decision about federation.

Architecture and Components

Query Processing. Search and Usage Analytics.

Query Processing

Query Processing Component analyzes and processes search queries and results.

When the query processing component receives a query, it performs linguistic processing first (like word breaking and stemming), then analyzes and further processes the query to optimize precision, recall, and relevance. In the end, the transformed query will be submitted to the index component.

The index component returns a result set based on the processed query, and the Query Processor in turn processes that result set, before returning it to the search front-end.

Search and Usage Analytics

The Search Engine's Analytics Processing Component performs both Search Analytics and Usage Analytics. This component uses information from these analyses to improve search relevance, create search reports, and generate recommendations and deep links.

Architecture and Components

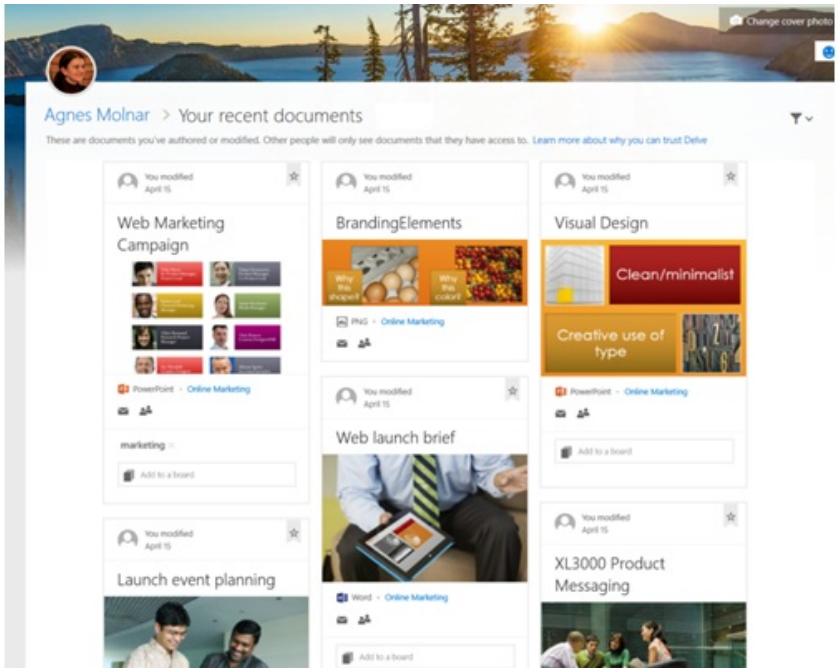
Office Graph. Delve.

Office Graph

Office Graph is a back-end service in Office 365 which uses machine learning techniques to provide intelligent discovery features for the users. It stores information about persons, content, and relationships between them.

When Hybrid Search is configured, even on-premises documents can be stored in Office Graph.

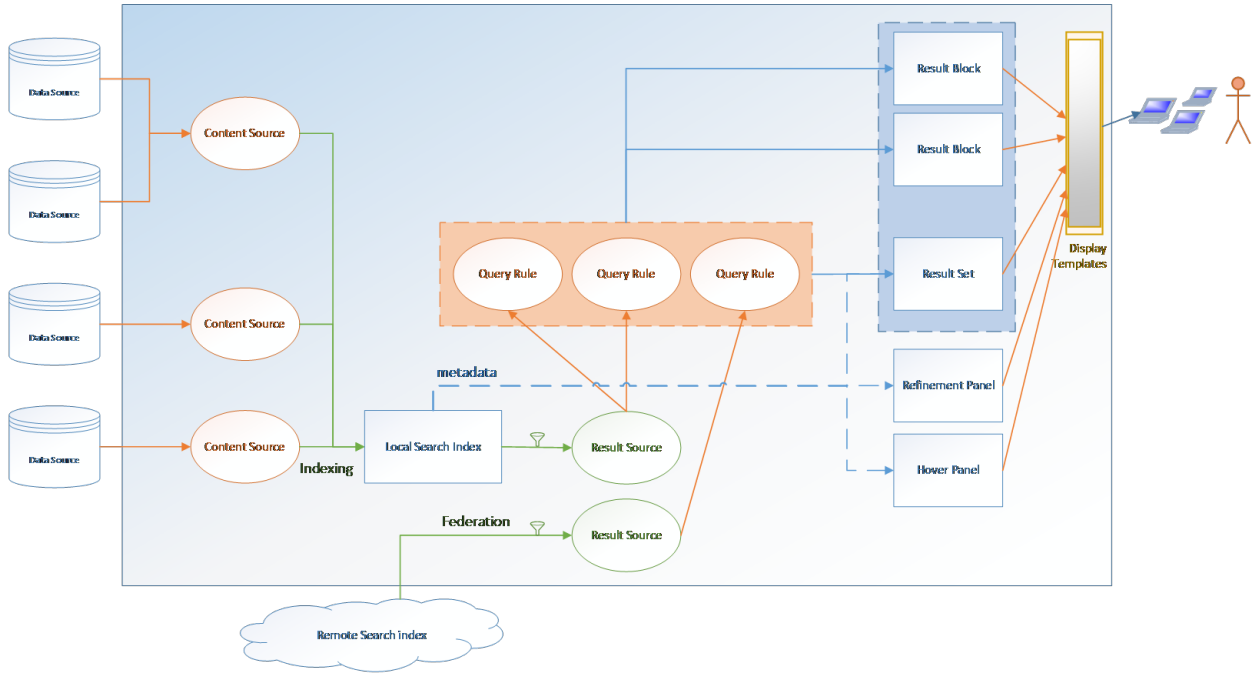
Office Delve



Office Delve is Office 365's intelligent discovery tool, which provides several views to support exploring content and other users. Delve is powered by Office Graph.

Search Features

SharePoint and Office 365 Search Features



Content

Content Source. Result Source. Content Freshness.

Content Source

In SharePoint, we define the connections between Search and a Source System by setting up a Content Source. Content Source definition contains the connection details to the Source System, as well as the crawl schedule settings.

Content Sources can be defined in the Search Service Application.

Result Source

In SharePoint and Office 365, Result Sources can be used to create sub-segments of the entire index, defined by property-based filters.

Some common examples: customer documents, tasks, expired cases, etc.

On the User Interface, Result Sets can be customized to limit the display of results only to a specific Result Source.

Content Freshness

Content Freshness measures the average time between a content is created or modified and its indexing.The shorter time passes from the creation/update, the better Content Freshness we have.

User Experience

Query. Search Results. Result Types.

Query

Query is the word or expression which initiates the Searching action by a user or an automated process. The Query is analyzed and processed by the Query Processing Component, then goes to the Indexer Component. The Indexer Component then gives back the proper and relevant, security trimmed results, and then these are displayed to the user who initiated the Query.

Simply said, Query is the question, while Result Set is the Search Engine's answer.

Search Results / Result Set

Search Results are the Search Engine's response to the Query. In SharePoint and Office 365, Search Results are always security trimmed, which means users can only see those items in the Search Result Set, which they have at least read access to.

Result Types

When displaying Search Results, it is important to show the relevant properties which help the searching user to find the most relevant items in the shortest possible time. However, these properties can vary from type to type: for example, contracts have different metadata than meeting notes, or marketing presentations.

In SharePoint and Office 365, Result Types provide elegant solution for this. With the help of them, we can define how each type of result should be displayed on the Result Set.

User Experience

Query Rules. Best Bets. Faceted Search / Refiners.

Query Rules

Query Rules are used for transforming the query by a pre-defined rule set. Query Rules are very useful to support user intent.

With Query Rules, we can display promoted results (Best Bets) in separate Result Blocks; can transform the query; or can define keyword-driven transformations.

Best Bets

Best Bets are promoted search results. Their primary goal can be to catch the users' attention, advertise some content, or promote items which are important to know about.

Best Bets are configured via Query Rules in SharePoint 2013, SharePoint 2016, and Office 365.

Faceted Search / Refiners

Search Facets (in standard Search terminology), or Refiners (in SharePoint / Office 365 language) are pre-configured filtering options, which are displayed after the Result Set is compiled and displayed.

Refiners are usually displayed on the sidebar, next to the Result Set. Their values depend on the current Result Set, therefore, may vary from Query to Query, moreover from User to User.

User Experience

Hover Panel. Display Template.

Hover Panel

Hover Panel is a dynamic panel over each Search Result, which gets displayed when the user hovers the mouse over an item on the Result Set.

Hover Panel shows the document preview for the most important file formats (Office documents, web pages, etc.), properties of the result, and actions to take. What and how is displayed on the Hover Panel can be defined by Display Templates.

Display Template

In SharePoint 2013, SharePoint 2016, and Office 365, User Interface elements, such as Result Set, Refiners, Hover Panel, are displayed in a new, customizable way. The customization can be done via HTML and JavaScript, enclosed into Display Templates.

Display Templates can be defined on the Site Collection level, and are stored in the MasterPage Gallery.

Metadata

Content Metadata. Search Metadata.
Crawled Property. Managed Property.

Content Metadata

Content Metadata is the set of properties, which the content has in the Source System. In most cases, different systems (and subsystems) have different set of properties, therefore unifying them is a critical step in every Enterprise Search implementation.

Search Metadata

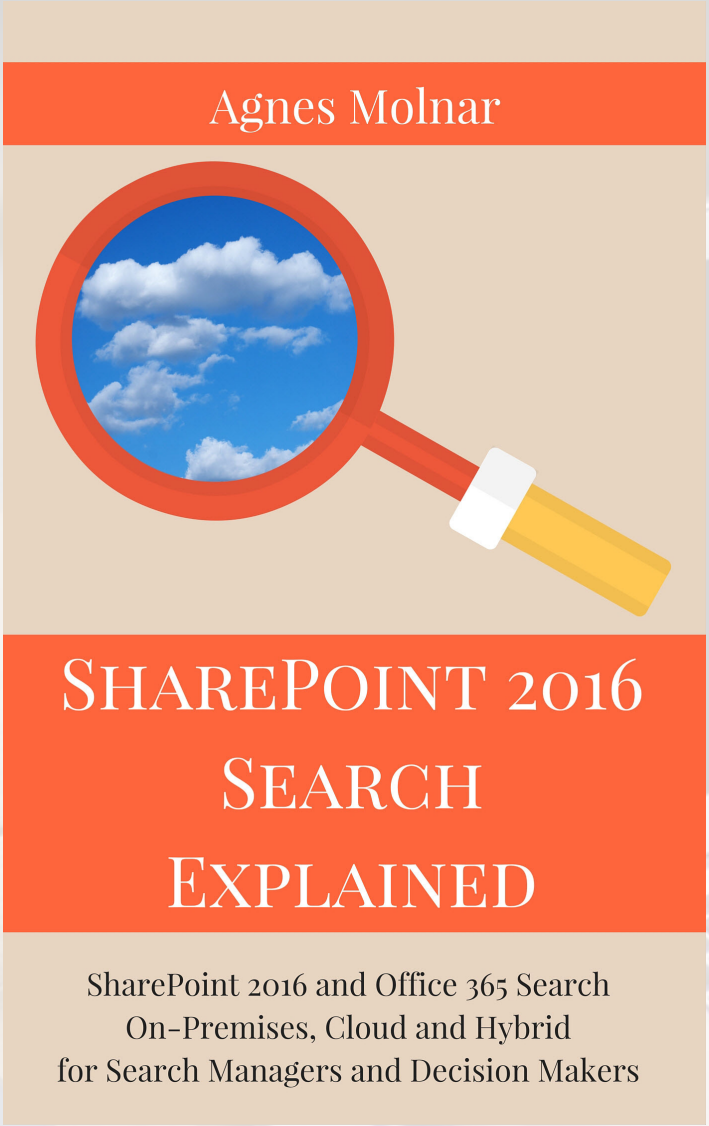
Search Metadata is defined in the Search Schema, and consists of Crawled Properties and Managed Properties (see below). Search Metadata is essential for every Search Application, as it describes the results, can be used for filtering and sorting the results, refiners, and also display on the Result Set or Hover Panel.

Crawled Property

Crawled Properties are the representations of the Content Metadata in SharePoint and Office 365 Search Schema.

Managed Property

Managed Properties are the Search Properties which can be used on the User Interface, and in Search Applications. To define where to get their values from, they have to be mapped to the proper Crawled Properties.



SharePoint 2016 Search Explained

Book and e-book

for Search Managers and Decision Makers

Buy NOW!

Contact

Question? Feedback?
If you have any questions, comments or feedback, please don’t hesitate to let me know!

E-mail: Agnes.Molnar@SearchExplained.com