



Lebanese University

Faculty of Information

Branch II

Dr. Mohamad Nagi

PhD in Data Mining - Computer Science

Metadata

Data Science - Semester 4

› Outlines

- › Metadata : Introduction, Definition & Concept
- › Need, Purpose and Functions of Metadata
- › Types of Metadata
- › Metadata Elements & Standards/ Schema
- › Dublin Core : Introduction, Background
- › Level of Standards & Elements

Outlines

- › Metadata & SEO
- › Metadata Editor
- › Metadata Extraction
- › Tags

Introduction

- › Metadata is key to the functionality of the systems holding the content, enabling users to find items of interest, record essential information about them, and share that information with others.
- › The cultural heritage world—libraries, archives, and a long history of creating and sharing robust, structured museums— has metadata.
- › For libraries, this takes the form of the library catalog, Early library catalogs were merely large inventory books, which then were replaced by catalog cards in drawers.
- › With computerization, libraries first moved to dedicated search terminals, and then in the Internet era to today's Web-based resource discovery systems.

METADATA : Definition & Concept

- › Metadata is Data about data.
- › Metadata is a set of data that describes and gives information about other data.
- › “Metadata is a structured information that describe, explains, locates or otherwise make it easier to retrieve , use and manage an information resource”.
- › Metadata describes how and when and by whom a particular set of data was collected, and how the data is formatted.
- › Metadata is defined as the data providing information about one or more aspects of the data; it is used to summarize basic information about data which can make tracking and working with specific data easier.

METADATA : Concept

- › Metadata makes it much easier for someone to locate a specific document.
- › For example, **Author**, **Date created** and **date modified** and **file size** are examples of very basic document metadata.
- › An image may include metadata that describes how large the picture is, the color depth, the image resolution, when the image was created, and other data.
- › A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.

METADATA : Concept

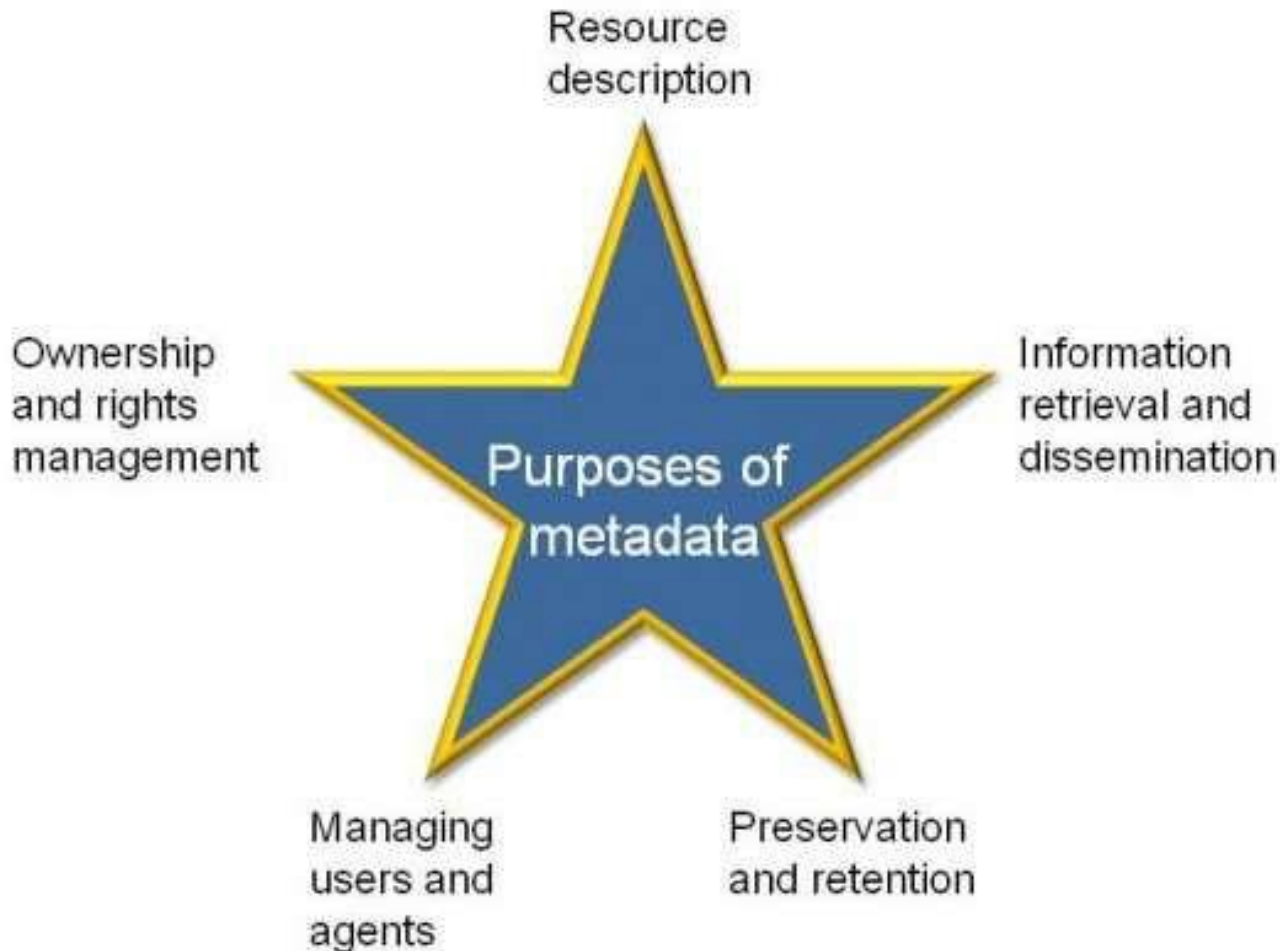
- › Web pages often include metadata in the form of meta tags. Description and keywords meta tags are commonly used to describe the Web page's content.
- › Metadata can be created manually, or by automated information processing.
- › Manual creation tends to be more accurate, allowing the user to input any information they feel is relevant or needed to help describe the file.
- › Automated metadata creation can be much more elementary, usually only displaying information such as file size, file extension, when the file was created and who created the file.

Why do we need Metadata?

Metadata is created and collected because it enables and improves use of that data. Metadata ensures that we will be able **find data, use data, and preserve and re-use data in the future.**

- › To enable discovery of your digitized material
- › To help you organize your digitized material
- › To support archiving and preservation

Purposes of Metadata



Functions of Metadata

- Metadata serves following functions :

› Resource discovery

- › Allowing resources to be found by relevant criteria;
- › Identifying resources;
- › Bringing similar resources together;
- › Distinguishing dissimilar resources;
- › Giving location information.

Functions of Metadata

- › **Organizing e-resources**
- › Organizing links to resources based on audience or topic.
- › Building these pages dynamically from metadata stored in databases.

Functions of Metadata

› Facilitating interoperability

- › Using defined metadata schemes, shared transfer protocols, and crosswalks between schemes, resources across the network can be searched more seamlessly.
 - Cross system search, e.g. using z39.50 protocol
 - Metadata harvesting e.g. using OAI protocol

Functions of Metadata

› Digital identification

- › Elements for standard number, e.g. ISBN.
- › The location of digital object may also given using :
 - A file name
 - A URL

Functions of Metadata

› Archiving and preservation

› Challenges :

- Digital information is fragile and can be corrupted or altered;
- It may become unusable as storage technologies change.

› Metadata is key to ensuring that resources will survive and continue to be accessible into the future. Archiving and preservation require special elements:

- to track the lineage of a digital object,
- to detail its physical characteristics
- to document its behavior in order to emulate it in future technologies.

Types of Metadata

- › The metadata application is manifold, covering a large variety of fields, there are specialized and well-accepted models to specify types of metadata.
- › **Bretherton & Singley** (1994) distinguish between two distinct classes: **structural/control** metadata and **guide** metadata.
 - › **Structural metadata** describes the structure of database objects such as tables, columns, keys and indexes.
 - › **Guide metadata** helps humans find specific items and are usually expressed as a set of keywords in a natural language.

Types of Metadata

- › **Ralph Kimball** divided metadata into 2 similar categories:
 - › **Technical metadata** corresponds to internal metadata,
 - › **Business metadata** corresponds to external metadata.
- **Kimball** added a third category, **process metadata**.

Types of Metadata

- › **NISO** National Information Standards Organization distinguishes among three types of metadata:

Descriptive metadata, Structural metadata, and Administrative metadata.

- › ***Descriptive metadata*** is typically used for discovery and identification, as information to search and locate an object, such as title, author, subjects, keywords, publisher.
- › ***Structural metadata*** describes how the components of an object are organized. An example of structural metadata would be how pages are ordered to form chapters of a book.

Types of Metadata

- › **Administrative metadata** gives information to help manage the source. Administrative metadata refers to the technical information, including file type, or when and how the file was created and who can access it.
- › There are several subsets of administrative data; two that are sometimes listed as separate metadata types are:
 - › **Rights management metadata**, which deals with intellectual property rights,
 - › **Preservation metadata**, which contains information needed to archive and preserve a resource.

Types	Definition	Examples
Administrative	Metadata used in managing and administering collections and information resources.	▪ Acquisition information ▪ Rights and reproduction tracking ▪ Documentation of legal access requirements ▪ Location information
Descriptive	Metadata used to identify and describe collections and related information resources.	▪ Cataloguing records ▪ Finding aids ▪ Differentiations between versions ▪ Hyperlinked relationships between resources
Preservation	Metadata related to the preservation management of collections and information resources	▪ Documentation of physical condition of resources ▪ Documentation of actions taken to preserve physical and digital versions of resources. ▪ Documentation of any changes occurring during digitization or preservation.

Types	Definition	Example
Technical	Metadata related to how a system functions or metadata behaves.	▪ Hardware and software documentation ▪ Technical digitization information ▪ Tracking of system response times ▪ Authentication and security data, e.g., encryption keys, passwords.
Use	Metadata related to the level and type of use of collections and information resources.	▪ Circulation records ▪ Physical and digital exhibition records ▪ Use and user tracking ▪ Content reuse and multi versioning information ▪ Search logs ▪ Rights metadata

Metadata Elements

- › The following are recommended as a minimum set of metadata elements. It is important to select a metadata standard or schema and consult that schema for complete information on each element. You may choose to use more elements based on the needs of your project.
- › **Title/Name** – Name given to the resource.
- › **Description** – A description of the resource and its spatial, temporal or subject coverage.
- › **Format** – File format, physical medium, dimensions of the resource, or hardware and software needed to access the data.

Metadata Elements

- › **Metadata** – Description of the metadata to be provided along with the generated data and a discussion of the metadata standards used, including the version of the schema and where the schema can be found.
- › **Identifier** – A unique identification assigned to the resource.
- › **Rights Holder** – The entities or persons who hold the rights to the data.
- › **Rights** – Information about the rights held in and over the resource.
- › **Contact Information** – Identity of, and means to communicate with persons or entities associated with the data.

Metadata Standards/Schema

- › In order to be useful, metadata needs to be **standardized**. This includes agreeing on language, spelling, date format, etc. If everyone uses a different standard, it can be very difficult to compare data to other data.
- › A key component of metadata is the **schema**. Metadata schemes are the overall structure for the metadata. It describes how the metadata is set up, and usually addresses standards for common components of metadata like dates, names, and places. There are also discipline-specific schemas used to address specific elements needed by a discipline.

Metadata Standards/Schema

General purpose schema

› Dublin Core

Dublin Core is a general standard first used by libraries, and can be adapted for specific disciplines. Dryad (www.datadryad.org), a digital data repository, uses Dublin Core.

› MODS (Metadata Object Description Schema)

This descriptive metadata schema is richer than Dublin Core, and can be used on its own or as a complement to other metadata formats.

Metadata Standards/Schema

Science Schema

› Darwin Core

This metadata schema is for describing biological specimens, including their occurrence in nature as documented by observations, samples, and related information. Based on Dublin Core, this schema is used in natural history specimen collections and species observation databases.

› Ecological Metadata Language (EML)

This metadata schema is for ecological data. EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data.

DUBLIN CORE Introduction

- › Finding relevant information on the World Wide Web has become increasingly problematic due to the explosive growth of networked resources. Current Web indexing evolved rapidly to fill the demand for resource discovery tools, but that indexing, while useful, is a poor substitute for richer varieties of resource description.
- › An invitational workshop held in March of 1995 brought together librarians, digital library researchers, and text-markup specialists to address the problem of resource discovery for networked resources.

What is DUBLIN CORE

- › The **Dublin Core Schema** is a small set of vocabulary terms that can be used to describe web resources (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like art works.
- › **Dublin Core** is an initiative to create a digital "library card catalogue" for the Web.
- › **Dublin Core** is made up of 15 **metadata** (data that describes data) elements that offer expanded cataloguing information and improved document indexing for search engine programs.
- › The Dublin Core Metadata Element Set is a general-purpose scheme for resource description originally intended to facilitate discovery of information objects on the Web.

What is DUBLIN CORE

- › The development of official specifications related to the Dublin Core is managed by the Dublin Core Metadata Initiative (DCMI), which consists of a small, paid directorate advised by a board of trustees, and a large number of loosely organized volunteers.
- › Dublin Core Metadata may be used for multiple purposes, from simple resource description, to combining metadata vocabularies of different metadata standards, to providing interoperability for metadata vocabularies in the Linked Data cloud and Semantic Web implementations.
- › The Dublin Core Metadata Element Set is a standard for cross domain resource description.

DUBLIN CORE Background

- › "Dublin" refers to Dublin, Ohio, USA where the schema originated during the 1995 invitational OCLC/NCSA Metadata Workshop, hosted by the Online Computer Library Centre (OCLC), a library consortium based in Dublin, and the National Centre for Supercomputing Applications (NCSA).
- › "Core" refers to the metadata terms as "broad and generic being usable for describing a wide range of resources".
- › The semantics of Dublin Core were established and are maintained by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, museums, and other related fields of scholarship and practice.

DUBLIN CORE Metadata Elements

A set 18 elements designed to enhance discovery and retrieval of resources.

Goals of DCME

- › Simplicity of creation and maintenance
- › Commonly understood semantics
- › Conformance to existing and emerging standards
- › International scope and applicability
- › Extensibility
- › Interoperability among collections and indexing system.

Why Use DUBLIN CORE

- › “The scope of Dublin core is specially designed to provide a metadata vocabulary of core properties able to provide basic description about any kind of resources... regardless of any format of media specialization or cultural origin. It is important that a semantic model used for resource discovery is not dependent on the medium of the source it means to describe...”
- › The Dublin Core metadata vocabulary is the result of many years of collaborative research to determine a common set of properties universal for describing any type of resource. The use of a standardized general classifications system also enables metadata of such collections to be combined and for knowledge contained within each collection to be shared.

DUBLIN CORE Level of Standards

The Dublin Core standard originally includes two levels: Simple and Qualified.

- › Simple Dublin Core comprised 15 elements i.e. Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format Identifier, Source, Language, Relation, Coverage, Right.
- › Qualified Dublin Core included 3 additional elements i.e. Audience, Provenance and Rights Holder.

DUBLIN CORE Simple vs Qualified

- › "Simple Dublin Core" is Dublin Core metadata that uses no qualifiers; only the main 15 elements of the Dublin Core Metadata Element Set are expressed as simple attribute-value pairs without any "qualifiers" (such as encoding schemes, enumerated lists of values, or other processing clues) to provide more detailed information about a resource.
- › "Qualified Dublin Core" employs additional qualifiers to further refine the meaning of a resource. One use for such qualifiers are to indicate if a metadata value is a compound or structured value, rather than just a string.
- › Qualifiers allow applications to increase the specificity or precision of the metadata. They may also introduce complexity that could impair the metadata's compatibility with other Dublin Core software applications.
- › A "date" is one example of a DC element that has the option of being further specified to identify it as a particular kind of date (date last modified, date published, etc.).

DUBLIN CORE Elements

1. Identifier: Title

Definition: A name given to the resource.

2. Identifier: Creator

Definition: An entity primarily responsible for making the content of the resource.

3. Identifier: Subject

Definition: The topic of the content of the resource.

4. Identifier: Description

Definition: An account of the content of the resource.

DUBLIN CORE Elements

5. Identifier: Publisher

Definition: An entity responsible for making the resource available.

6. Identifier: Contributor

Definition: An entity responsible for making contributions to the content of the resource.

7. Identifier: Date

Definition: A date associated with an event in the life cycle of the resource.

8. Identifier: Type

Definition: The nature or genre of the content of the resource.

DUBLIN CORE Elements

9. Identifier: Format

Definition: The physical or digital manifestation of the resource.

10. Identifier: Identifier

Definition: An unambiguous reference to the resource within a given context.

11. Identifier: Source

Definition: A reference to a resource from which the present resource is derived.

12. Identifier: Language

Definition: A language of the intellectual content of the resource.

DUBLIN CORE Elements

13. Identifier: Relation

Definition: A reference to a related resource.

14. Identifier: Coverage

Definition: The extent or scope of the content of the resource.

15. Identifier: Rights

Definition: Information about rights held in and over the resource.

16. Identifier : Audience

Definition : A class of entity for whom the the resource is intended or useful.

DUBLIN CORE Elements

17. Identifier: Provenance

Definition : A statement of any change in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation.

18. Identifier : Right Holder

Definition : A person or organization owning and managing rights over the resource.

DUBLIN CORE

References

- Kunze, J. and T. Baker, “The Dublin core metadata elements set”,2013.
- Baker Thomas, “A Grammar of Dublin Core” ,2011.
- <http://marciazeng.slis.kent.edu/metadatabasics/types.htm>. Retrieved on April 12, 2017.
- <http://www.kcoyle.net/jal-31-2.html> . Retrieved on April 12,2017.
- <http://dublincore.org>. Retrieved on April 18, 2017.
- http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Retrieved on April 16, 2017.
- <http://www.loc.gov/standards/metadata.html#types>. Retrieved on April 16, 2017.
- <https://www.practicalecommerce.com/SEO-Why-Is-Metadata-Important>

DUBLIN CORE

There is two Dublin Core online Generators:

- Simple Dublin core generator:

https://nsteffel.github.io/dublin_core_generator/generator_nq.html#date

- Advanced Dublin Core Generator:

https://nsteffel.github.io/dublin_core_generator/generator.html

Metadata And SEO

It's critical that ecommerce marketers understand the metadata that drives search engine optimization.

Metadata is a series of micro-communications between your site and search engines.

Nearly all metadata is invisible to visitors. It lives and works behind the scenes in the HTML of web pages. The metadata we use for SEO speaks to search engines directly from each page crawled. Because it's not immediately visible, metadata can seem foreign. Here is a typical example, found in nearly all pages.

```
<title>The title tag goes here</title> <meta name="description"
content="And the descriptive text that goes in here is the meta
description. "> <meta name="keywords" content=""/>
```

Metadata And SEO

- SEO (Search Engine Optimisation) involves altering website code, content, and presence in order to increase rankings in search engines
- At least half a billion searches are made worldwide on a daily basis
 - If you are ranked at the top of a search engine, you are essentially running a free marketing campaign

Metadata And SEO

How a search engine works?

- Search engines use programs called spiders, or bots, to search the internet and document their findings.
- Spiders firstly go to a webpage and analyse it
 - Goes through the HTML code and categorises the most important words (keywords) in the document.
- It then moves on to any pages that linked from the current page.
 - This is one way how search engines find new pages.
 - It also visits pages that may not be linked by other pages, but were submitted by search engines.

Metadata And SEO

Search Engine Optimization:

- SEO consists of activities / techniques applied towards making a website rank at the top in search engine.
- The prime aim of SEO exercise for a marketer is to help generate awareness for its brand website by gaining attention through **content relevance** and **keyword prominence** rather than paying for it through SEM campaigns.
- The key aim on any search engine is to help users find a specific piece of information from the billion pieces of content lying on the internet.
- Search Engines execute following **major functions which** form the core of their businesses.



Metadata And SEO

Crawling:

The purpose of the crawl is to **identify relevant pages for indexing** and **assess whether they have changed**.

Crawling is performed by **robots (bots)** which are also known **as spiders**. These access web pages and retrieve a reference URL of the page for later analysis.

Metadata And SEO

Indexing:

An index is created to **enable the search engine to rapidly find the most relevant pages containing the query typed by the searcher.**

Rather than searching each page for a query phrase, a search engine '**inverts**' the index to produce a lookup table of the documents containing particular words.

Metadata And SEO

Ranking & Scoring:

The indexing process has produced ***a lookup of all the pages*** that contain particular words in a query, but they are not sorted in terms of relevance.

Ranking of the document to assess the most relevant set of documents to return in the Search engine result pages (SERPs) occurs in real time for the search query entered.

First, relevant documents will be retrieved from a runtime version of the index at a particular data center, then a rank in the SERPs for each.

Metadata And SEO

Query request and Results Serving:

The familiar search engine interface **accepts** the searcher's query.

The **user's location** is assessed through their **IP address** and the query is then passed to a relevant data center for processing.

Ranking then occurs in real time for a particular query to return a sorted list of relevant documents, and these are then displayed **on the Search Results Page**.

Metadata And SEO

Improving Search Engine Ranking Through SEO

- The challenge for SEOs is that there are many **ranking factors** and they are constantly changing!
- Google references over **200 main ranking factors** which mean hundreds of detailed changes to the way pages are evaluated and represented in the search results places happening each year.
- Fortunately there are **four common factors** that influence search engine rankings, that can be managed as part of an SEO strategy:

Metadata And SEO

Important SEO Tips You Need to Know:

1. Remove anything that slows down your site.
2. Link to other websites with relevant content.
3. Write for humans first, search engines second
4. Encourage other trustworthy sites to link to you.
5. Write unique and relevant Meta descriptions for every page.
6. Use readable and meaningful URLs only.
7. Use the right keywords in your images.
8. Create & publish unique content consistently to improve your rankings.
9. Build momentum with social signals.

Metadata And SEO

Meta Tags

Pay careful attention to the terminology. Meta tags are metadata, but not all metadata are meta tags. Some elements commonly called “tags” are actually attributes of a tag.

Meta Tags

The most obvious metadata for SEO are meta tags, so we'll start there. The meta tag takes the following form.

```
<meta name="description" content="This what a meta tag with a  
name attribute of description looks like." />
```

Each of the tags below follows this format with the beginning tag of “meta,” followed by name attributes. When you type a meta description into your CMS, it automatically generates the meta tag in the correct format.

Metadata And SEO Description

Description: Sometimes used by search engines as the descriptive black text in the search result listing, meta descriptions can help increase customer clicks in search results, but meta descriptions will not impact rankings. The description attribute for the meta tag explains the page content in a summary that needs to be at least 11 words long to display, and will truncate at around 160 characters. (After roughly 160 characters, the remaining text won't display.) Don't bother placing the first 160 characters from the copy on the page into your meta description — just leave it blank if you have to. Google will ignore such descriptions and determine the most relevant content (from the text on the page) to display as the summary text in its search results.

Metadata And SEO Keywords

Keywords:

Meta keywords ceased to impact rankings in 2009.

Bing may still use them, but only as a spam signal. In other words, too many irrelevant keywords in this attribute may harm rankings in Bing. Do not use the keyword attribute unless your internal site search engine requires it.

Metadata And SEO Robots

Robots:

Part of the exclusion protocol, the meta robots attribute tells search engines whether to index or pass link authority through the links on a page. The four attributes are “index,” “noindex,” “follow,” “nofollow.” Keep in mind that search engines by default index content and follow links. So it’s pointless to use the attribute combination of “index, follow.”

Meta robots attributes are commonly used at the template level, since you may want to exclude all pages using a certain template from being indexed. Remember, though, that accidental use of the *noindex* tag can result in drastic decreases in SEO performance.

Metadata And SEO

Title Tags

Title Tags

Title tags are still the single most important piece of metadata on the page. Their format is simple, and as with meta tags, your CMS will generate the tag for you from the title or headline you enter. Here's what a title tag looks like.

```
<title>SEO: For Conversions, Every Page Is a Landing Page |  
Practical Ecommerce</title>
```

The best title tags begin with the most relevant keywords, product name, or article name, and end with the name of the site — “Practical Ecommerce,” above. Stay within 60 characters and keep the unique, relevant, and valuable keywords toward the beginning of the title for maximum SEO benefit.

Metadata And SEO

Schema Markup

Schema Markup: In 2011, Google, Bing, and Yahoo launched the Schema.org project to enable webmasters to mark their pages with a specific syntax of data to help the engines digest the content more accurately and efficiently. In particular, the Schema markup about your company, store, and your products' prices, availability, and ratings are instrumental in allowing the search engines to display information about your items directly in the search results.

Unlike the other metadata listed here, Schema structured data is code that needs to be implemented by a developer. It is most commonly seen as a series of JSON-LD instructions that look like this:

```
<script type="application/ld+json"> { "@context":  
"http://schema.org", "@type": "Organization", "url":  
"http://www.example.com", "name": "ACME Sales  
Corp.", "contactPoint": { "@type": "ContactPoint", "telephone": "+1-  
888-888-7890", "contactType": "Customer service" } } </script>
```

Metadata And SEO Image Tags

Image Tags and Alt Attributes

Search engines primarily use image tags to identify the URLs for the images to be shown on the page. Other attributes can specify the height and width of the image. Here's what an image tag looks like.

```

```

Alt is the most important image attribute for SEO, though its relative importance to other SEO factors is small. Alternative attributes provide a short textual description of the image to serve accessibility needs for customers with vision or mobility disabilities. They also serve as small keyword relevance signals for SEO, most importantly in image search.

Metadata Editor

To add metadata to a website :

- › if the website is HTML we can use the Dublin core metadata generator to generate the DC Metadata and then we can export the Code as HTML and then add the code to the desired website in the head section.
- › If the website is built by an opensource engine like WordPress or any other CMS, we can install one or many tools as plugins to our website, which will increase the performance of the website and simplify the keywords and metadata retrieval by the search engine.

Metadata Editor

To add metadata to a file

There is many tools and programs, so we can use any of these to add the metadata.

For Photos and videos and other media files we can use fotostation software, adobe bridge and others..

To Download fotostation application:

<https://fotostation.com/download/>



Metadata Extraction

To Extract metadata from photos there is an online extractor

<http://exif.regex.info/exif.cgi>

Metadata Extraction

To extract Metadata and keywords from a website, many online extractor are present, for example we can use the following link to extract metadata, page title, page description and keywords from any website on the cloud:

<http://tools.buzzstream.com/meta-tag-extractor>

To Extract links from a specific website:

<http://tools.buzzstream.com/link-building-extract-urls>

Metadata and Tags

Metadata is one of those terms many of us might hear often and even use daily, yet not fully understand what it means. In short, metadata is data about data. It helps the search for relevant information and organizes electronic resources.

Additionally, metadata provides digital identification, and supports the archiving and preservation of electronic resources.

There are two types of metadata that you need to be aware of:

- The first is **structural metadata**, which data about the design and specifications of data structures. To put that in another way, it is the data about the functions, shapes, colors, and layouts of things that hold the content you see online.
- There is also **descriptive metadata**, which is data bout individual instances of data or data content. This is the data that fills data structures. This can be website copy, images, blog articles, etc.

Metadata and Tags

Metadata is a way of organizing information in a way that is simple and grouped in a logical manner. The reason that understanding and using metadata is important is that it's what search engines use to find your website. Properly tagging and describing your content is important to increase your search engine optimization and get found by potential customers!

With the advent of the digital age, metadata became a way to describe digital data. Metadata's function once again evolved with Web 2.0. While metadata exists in the same form, end users/consumers use hashtags (pound symbol/number sign) to find related content to consume. This has greatly democratized the distribution of information and content across the internet and has been revolutionary in the sense that now anyone can be a publisher/content creator and distributor.

Metadata and Tags

As digital consumers ourselves, we are very familiar with the hashtag (#), a type of metadata that describes groups of related content. Hashtags are chosen by the content creator, informally, and when many content creators use the same hashtags, groupings of similar content are formed. Twitter was the one of the first platforms to use this, with Instagram and now even Facebook incorporating hashtags within their social platforms ecosystem. Users find related content using these hashtags, and even create communities around these different groupings.

As more and more users use that hashtag, a user could browse other tweets/pictures on Twitter with that hashtag and share their concert experience with others.

As a business, utilizing hashtags appropriately will contribute positively for creating brand excitement and increasing your brand's reach. Because hashtags on social media platforms are informally used.

Metadadata and Tags

Tagging your content online with social media hashtags isn't an exact science and you're not going to be successful all the time. However, there are some general guidelines to using hashtags. It's usually best practice to add one hashtag to a post, two if you include a location as well, and three at the absolute maximum. Using more in your post can infuriate users, and work against you while building your brand presence on that particular platform. Additionally, you could get your business account suspended for over-hashtagging.

Twitter warns you that misuse or overuse of hashtags will lead to your account being filtered from search results (bad), or being suspended (worse).

Metadata and Tags

More than just using hashtags on social media, using metadata correctly within your website is also very good for SEO. Metadata has democratized the media business because anyone can be a publisher and distribute their content widely and for free! This leads to the long tail effect of content generation, where specialized and engage communities based on content emerge from these groupings. By incorporating detailed meta descriptions, tags in your blog posts, and couple that with a light and responsive website, you'll increase your organic search greatly.

Metadata and Tags

The difference between metadata and hashtag

The difference is that **metadata** is data that describes other data, serving as an informative label while **hashtag** is (internet) a metadata tag, signaled by a preceding hash sign (#), used to label content.

Metadata and Tags

XML a Metadata Language

The power of XML in a sense is in the fact that it is a meta-language a language for defining other languages. So given an XML schema you can define a language such as HTML. Or I can define a language for expressing customers and all its related objects (such as "orders" and "contract") for a CRM (customer relationship management) application.

So the general idea is that for a given object I define a set of metadata or tags. Then I can use any of those tags to find the object easier. For example, I can define tags for an online ad. Then an ad server such as Google or Facebook can push that ad onto the pages where a viewer who is searching or reading about something that relates to that tag.

Metadata and Tags

Web 2.0 and Tagging

Modern Web 2.0 sites often provide a "tag cloud", a space in which the tags are randomly listed such that the size of a tag (font) represents the frequency of use of that tag.

When I look at a tag cloud I can right away see the most popular topics and trends and easily drill down on that topic by simply selecting that tag.

Tagging may also enable websites to make suggestions to the viewers. If I am reading news, researching a topic, or shopping for a product or service online, then the website based on the common tags, may suggest related topics, products or services to me.

Metadata and Tags

Geotagging

An interesting application of tagging applied to photos is *geotagging*; that is, tagging a photo with the latitude and longitude of the location where the photo is taken. Flickr supports geotagging and in fact there is a large Flickr group with geotagged photos.

The geotags which are denoted by "geo:lat" and "geo:lon" are often machine generated as it is not easy for human to do this. You essentially need a GPS system for this purpose. In fact there now digital cameras with a GPS add-on to automatically geotag the photos that you take!

Geotagging has useful applications. For example, photos with geotags can be mapped to a digital map such as Google Earth. Geotagging doesn't have to apply to photos only. Any object that has a location attribute (an address, person,) can be geotagged.

Metadata and Tags

Tagging in Facebook

Facebook takes photo tagging to a new level by enabling *regional tagging* in images. The most obvious application of this is allowing users to tag a small square of a photo (with typically a user). So I can tag a group photo with the individual people who are in the photo. Then the viewer of that photo by moving the mouse over the photo, can see the tags (i.e. the names of the people tagged). Note that this photo tagging is *collaborative* and *social*. Depending on the permissions i can add a tag to an image posted by someone else. And if i am tagged in an image, i can remove my tag.

Furthermore, that photo shows up on the profile and photo pages of all the people in the photo who are tagged (assuming they have an active profile on Facebook). In a sense the image tagging in Facebook becomes a means of communication and messaging.

Metadata and Tags

Hashtags in Twitter

Another powerful and interesting application of tagging is in Twitter. You can use hashtags (#<tags>) with any tweet (as part of the 140 characters and appearing anywhere in the tweet) which essentially directs a tweet to a particular virtual folder, list or bucket.

For example, Twitter related sites such as www.twitscoop.com use these hashtags to identify hot trends. By selecting a hashtag in a tweet or in a tag cloud, you can get the latest list of tweets with that hashtag.

There is much more to be said on Twitter and its hashtags which I will cover in another post.

XML

What is XML

- › XML stands for eXtensible Markup Language.
- › A markup language is used to provide information about a document.
- › Tags are added to the document to provide the extra information.
- › HTML tags tell a browser how to display the document.
- › XML tags give a reader some idea what some of the data means.



XML

What is XML used for ?

- › XML documents are used to transfer data from one place to another often over the Internet.
- › XML subsets are designed for particular applications.
- › One is RSS (Rich Site Summary or Really Simple Syndication). It is used to send breaking news bulletins from one web site to another.
- › A number of fields have their own subsets. These include chemistry, mathematics, and books publishing.
- › Most of these subsets are registered with the W3Consortium and are available for anyone's use.



XML

Advantages of XML

- › XML is text (Unicode) based.
 - › Takes up less space.
 - › Can be transmitted efficiently.
- › One XML document can be displayed differently in different media.
 - › Html, video, CD, DVD,
 - › You only have to change the XML document in order to change all the rest.
- › XML documents can be modularized. Parts can be reused.



XML

Example of an HTML Document

```
<html>
  <head><title>Example</title></head>
  <body>
    <h1>This is an example of a page.</h1>
    <h2>Some information goes here.</h2>
  </body>
</html>
```



XML

Example of an XML Document

```
<?xml version="1.0"/>
<address>
  <name>Alice Lee</name>
  <email>alee@aol.com</email>
  <phone>212-346-1234</phone>
  <birthday>1985-03-22</birthday>
</address>
```



XML

Difference Between HTML and XML

- › HTML tags have a fixed meaning and browsers know what it is.
- › XML tags are different for different applications, and users know what they mean.
- › HTML tags are used for display.
- › XML tags are used to describe documents and data.



XML

XML Rules

- › Tags are enclosed in angle brackets.
- › Tags come in pairs with start-tags and end-tags.
- › Tags must be properly nested.
 - › `<name><email>...</name></email>` is not allowed.
 - › `<name><email>...</email><name>` is.
- › Tags that do not have end-tags must be terminated by a '/'.
 - › `<br />` is an html example.



XML

More XML Rules

- › Tags are case sensitive.
 - › **<address> is not the same as <Address>**
- › XML in any combination of cases is not allowed as part of a tag.
- › Tags may not contain '<' or '&'.
- › Tags follow Java naming conventions, except that a single colon and other characters are allowed. They must begin with a letter and may not contain white space.
- › Documents must have a single *root* tag that begins the document.

XML

Predifined Entity

Code	Symbole	Description
<	<	less than
>	>	greater than
&	&	ampersand
'	'	apostrophe
"	"	quotation mark

XML Encoding

- › XML (like Java) uses Unicode to encode characters.
- › Unicode comes in many flavors. The most common one used in the West is UTF-8.
- › UTF-8 is a variable length code. Characters are encoded in 1 byte, 2 bytes, or 4 bytes.
- › The first 128 characters in Unicode are ASCII.
- › In UTF-8, the numbers between 128 and 255 code for some of the more common characters used in western Europe, such as ã, á, å, or ç.
- › Two byte codes are used for some characters not listed in the first 256 and some Asian ideographs.
- › Four byte codes can handle any ideographs that are left.
- › Those using non-western languages should investigate other versions of Unicode.



XML

Well-Formed Documents

- › An XML document is said to be well-formed if it follows all the rules.
- › An XML parser is used to check that all the rules have been obeyed.
- › Recent browsers such as Internet Explorer 5 and Netscape 7 come with XML parsers.
- › Parsers are also available for free download over the Internet. One is Xerces, from the Apache open-source project.
- › Java 1.4 also supports an open-source parser.



XML

XML Example Revisited

```
<?xml version="1.0"/>
```

```
<address>
```

```
  <name>Alice Lee</name>
```

```
  <email>alee@aol.com</email>
```

```
  <phone>212-346-1234</phone>
```

```
  <birthday>1985-03-22</birthday>
```

```
</address>
```

- Markup for the data aids understanding of its purpose.
- A flat text file is not nearly so clear.

Alice Lee

alee@aol.com

212-346-1234

1985-03-22

- The last line looks like a date, but what is it for?



XML

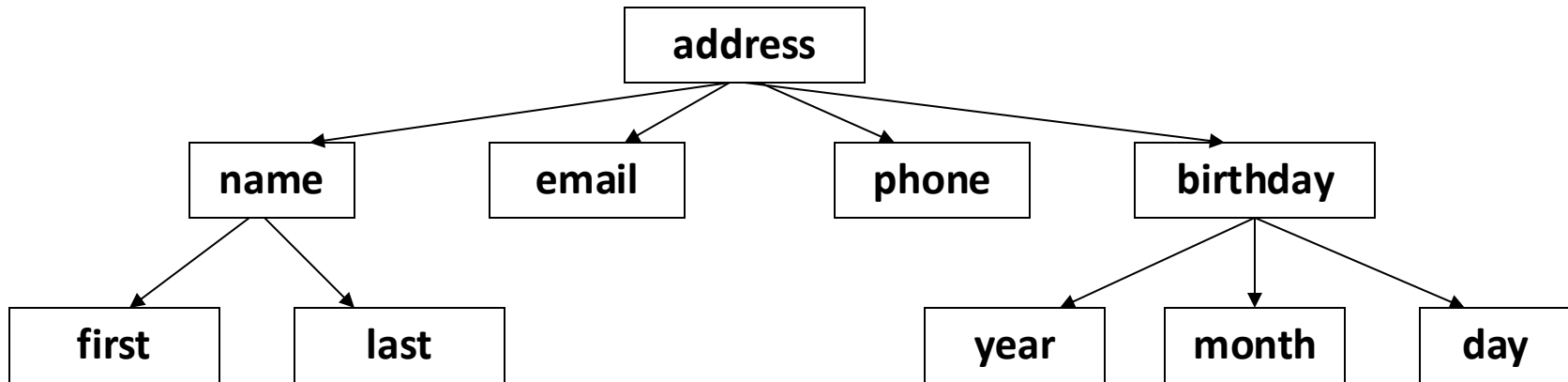
Expended Example

```
<?xml version = "1.0" ?>
<address>
  <name>
    <first>Alice</first>
    <last>Lee</last>
  </name>
  <email>alee@aol.com</email>
  <phone>123-45-6789</phone>
  <birthday>
    <year>1983</year>
    <month>07</month>
    <day>15</day>
  </birthday>
</address>
```



XML

XML Files are Trees



XML

XML Trees

- › An XML document has a single root node.
- › The tree is a general ordered tree.
 - › A parent node may have any number of children.
 - › Child nodes are ordered, and may have siblings.
- › Preorder traversals are usually used for getting information out of the tree.

XML

Validity

- › A well-formed document has a tree structure and obeys all the XML rules.
- › A particular application may add more rules in either a DTD (document type definition) or in a schema.
- › Many specialized DTDs and schemas have been created to describe particular areas.
- › These range from disseminating news bulletins (RSS) to chemical formulas.
- › DTDs were developed first, so they are not as comprehensive as schema.



XML

Document Type Definitions

- › A DTD describes the tree structure of a document and something about its data.
- › There are two data types, PCDATA and CDATA.
 - › PCDATA is parsed character data.
 - › CDATA is character data, not usually parsed.
- › A DTD determines how many times a node may appear, and how child nodes are ordered.



XML

DTD for address Example

```
<!ELEMENT address (name, email, phone, birthday)>  
<!ELEMENT name (first, last)>  
<!ELEMENT first (#PCDATA)>  
<!ELEMENT last (#PCDATA)>  
<!ELEMENT email (#PCDATA)>  
<!ELEMENT phone (#PCDATA)>  
<!ELEMENT birthday (year, month, day)>  
<!ELEMENT year (#PCDATA)>  
<!ELEMENT month (#PCDATA)>  
<!ELEMENT day (#PCDATA)>
```

XML Schemas

- › Schemas are themselves XML documents.
- › They were standardized after DTDs and provide more information about the document.
- › They have a number of data types including string, decimal, integer, boolean, date, and time.
- › They divide elements into simple and complex types.
- › They also determine the tree structure and how many children a node may have.



XML

Schema for First address Example

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="address">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="name" type="xs:string"/>
        <xs:element name="email" type="xs:string"/>
        <xs:element name="phone" type="xs:string"/>
        <xs:element name="birthday" type="xs:date"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```



XML

Explanation of Example Schema

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
```

- ISO-8859-1, Latin-1, is the same as UTF-8 in the first 128 characters.

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

- www.w3.org/2001/XMLSchema contains the schema standards.

```
<xs:element name="address">
```

```
  <xs:complexType>
```

- This states that address is a complex type element.

```
    <xs:sequence>
```

- This states that the following elements form a sequence and must come in the order shown.

```
<xs:element name="name" type="xs:string"/>
```

- This says that the element, name, must be a string.

```
<xs:element name="birthday" type="xs:date"/>
```

- This states that the element, birthday, is a date. Dates are always of the form yyyy-mm-dd.



XML

XSLT- Extensible Stylesheet Language Transformations

- › XSLT is used to transform one xml document into another, often an html document.
- › The Transform classes are now part of Java 1.4.
- › A program is used that takes as input one xml document and produces as output another.
- › If the resulting document is in html, it can be viewed by a web browser.
- › This is a good way to display xml data.



XML

XSLT- Extensible Stylesheet Language Transformations

A Style Sheet to Transform address.xml

```
<?xml version="1.0" encoding="ISO-8859-1"?>
  <xsl:stylesheet version="1.0"
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
    <xsl:template match="address">
      <html><head><title>Address Book</title></head>
      <body>
        <xsl:value-of select="name"/>
        <br/><xsl:value-of select="email"/>
        <br/><xsl:value-of select="phone"/>
        <br/><xsl:value-of select="birthday"/>
      </body>
    </html>
  </xsl:template>
</xsl:stylesheet>
```



XML

XSLT- Extensible Stylesheet Language Transformations

The Result of the Transformation

Alice Lee
alee@aol.com
123-45-6789
1983-7-15

XML Parsers

- › There are two principal models for parsers.
- › SAX – Simple API for XML
 - › Uses a call-back method
 - › Similar to javax listeners
- › DOM – Document Object Model
 - › Creates a parse tree
 - › Requires a tree traversal

References

- › Kunze, J. and T. Baker, “The Dublin core metadata elements set”,2013.
- › Baker Thomas, “A Grammar of Dublin Core” ,2011.
- › <http://marciazeng.slis.kent.edu/metadatabasics/types.htm>. Retrieved on April 12, 2017.
- › <http://www.kcoyle.net/jal-31-2.html> . Retrieved on April 12, 2017.
- › <http://dublincore.org>. Retrieved on April 18, 2017.
- › http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Retrieved on April 16, 2017.
- › <http://www.loc.gov/standards/metadata.html#types>. Retrieved on April 16, 2017.
- › Elliotte Rusty Harold, Processing XML with Java, Addison Wesley, 2002.
- › Elliotte Rusty Harold and Scott Means, XML Programming, O'Reilly & Associates, Inc., 2002.
- › W3Schools Online Web Tutorials, <http://www.w3schools.com>.