

Coveo Enterprise Search

How to Index Metadata with User Custom Fields

This document explains how to index metadata from various types of documents using Custom Fields.

What Is Metadata

The term metadata—or meta-information—refers to a type of data whose purpose is to provide information concerning other data in order to facilitate their management and understanding. The author, modification date and size of a document are examples of metadata. *Coveo Enterprise Search* (CES) indexes metadata from the following types of documents:

- *Microsoft Office* documents
- *Adobe* PDF files
- *SharePoint* list items
- Web pages
- *Microsoft Exchange* items
- *XML* documents

CES does not automatically extract user-defined metadata from your documents; therefore, it is not yet available for queries. You must first create a custom field for each type of metadata.

Indexing Metadata

The following steps describe how to index metadata from your documents:

1. Make sure that CES does not already extract the data through system or built-in custom fields. Refer to [What Is the Difference between Built-In and Custom Fields](#).
2. Identify the user-defined metadata information in your documents: [Metadata in Web Documents](#), [Metadata in Microsoft Office Documents](#), [Metadata in SharePoint list items](#) and [Metadata in SharePoint list items](#).
3. Access **Index > Sources and Collections > Custom Fields** to create a custom field with this information.
4. Rebuild the source.

Here is an overview of the required information:

- **Name:** Specify a name for the custom field. It will then be used in field queries like `@fieldname=fieldvalue`.
- **Type:** Metadata is always extracted from documents as text. Conversion is made afterwards, depending on this parameter and then inserted in the index.
- **Metadata Name:** Name of the metafield in the document itself.

It is now possible to perform advanced queries (ex.: `@docid=7`) with the new custom field.

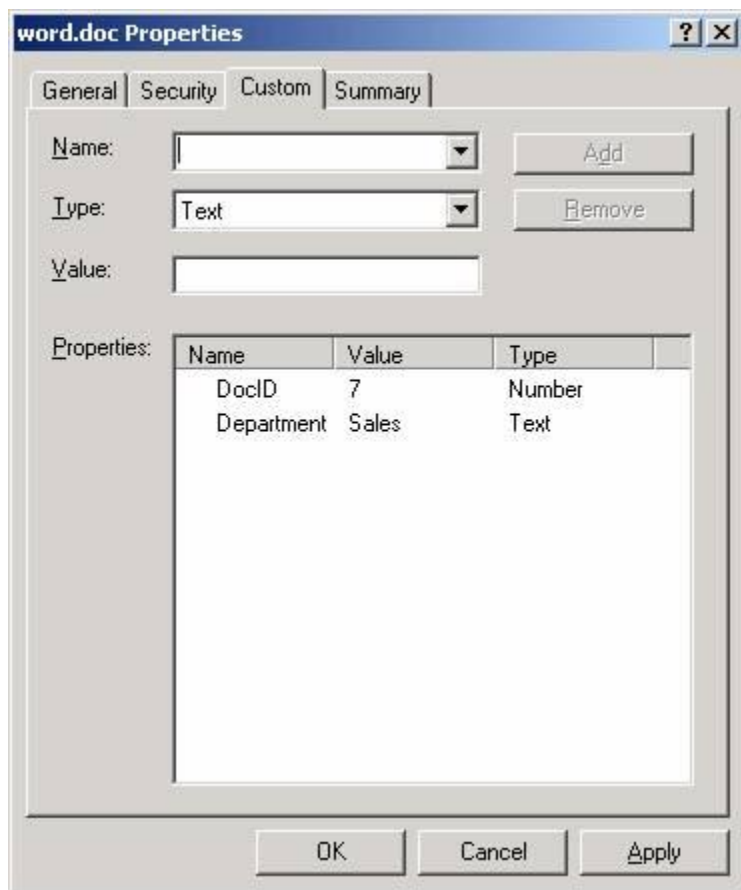
Metadata in Web Documents

Metadata information in a Web document can be found between the HTML **META** tags. For example, company XYZ adds two specific pieces of information in their documents: the department name and document internal identifier.

```
<html>
  <head>
    <meta name="Department" content="Sales">
    <meta name="DocId" content="7">
  </head>
  <body>
  </body>
</html>
```

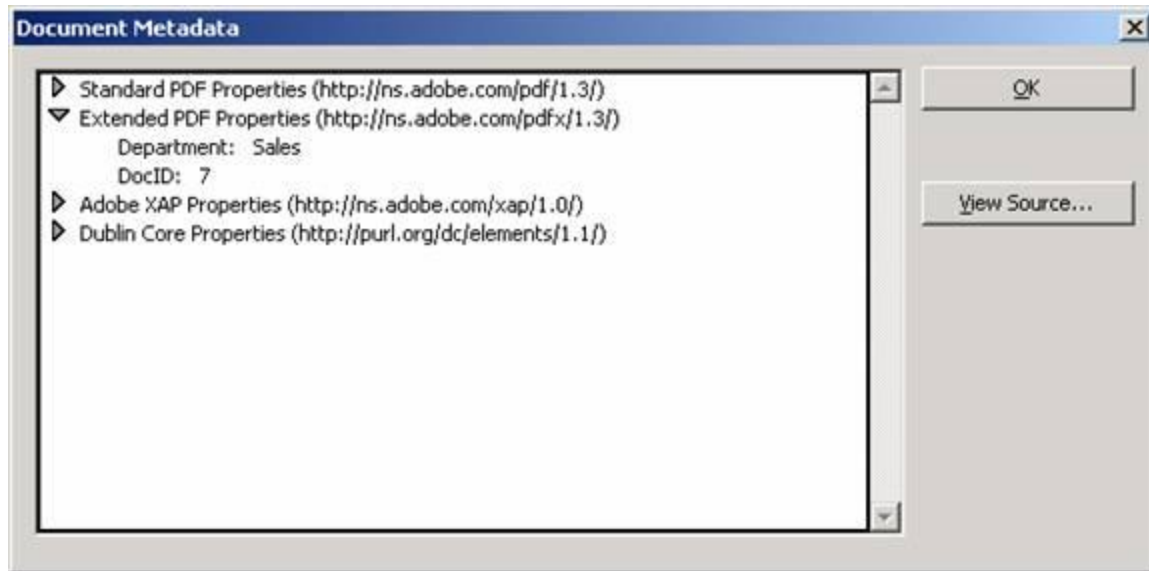
Metadata in Microsoft Office Documents

It is possible to view, add or modify metadata in *Microsoft Office* documents. Click **Properties** on the shortcut menu and select the **Custom** tab.



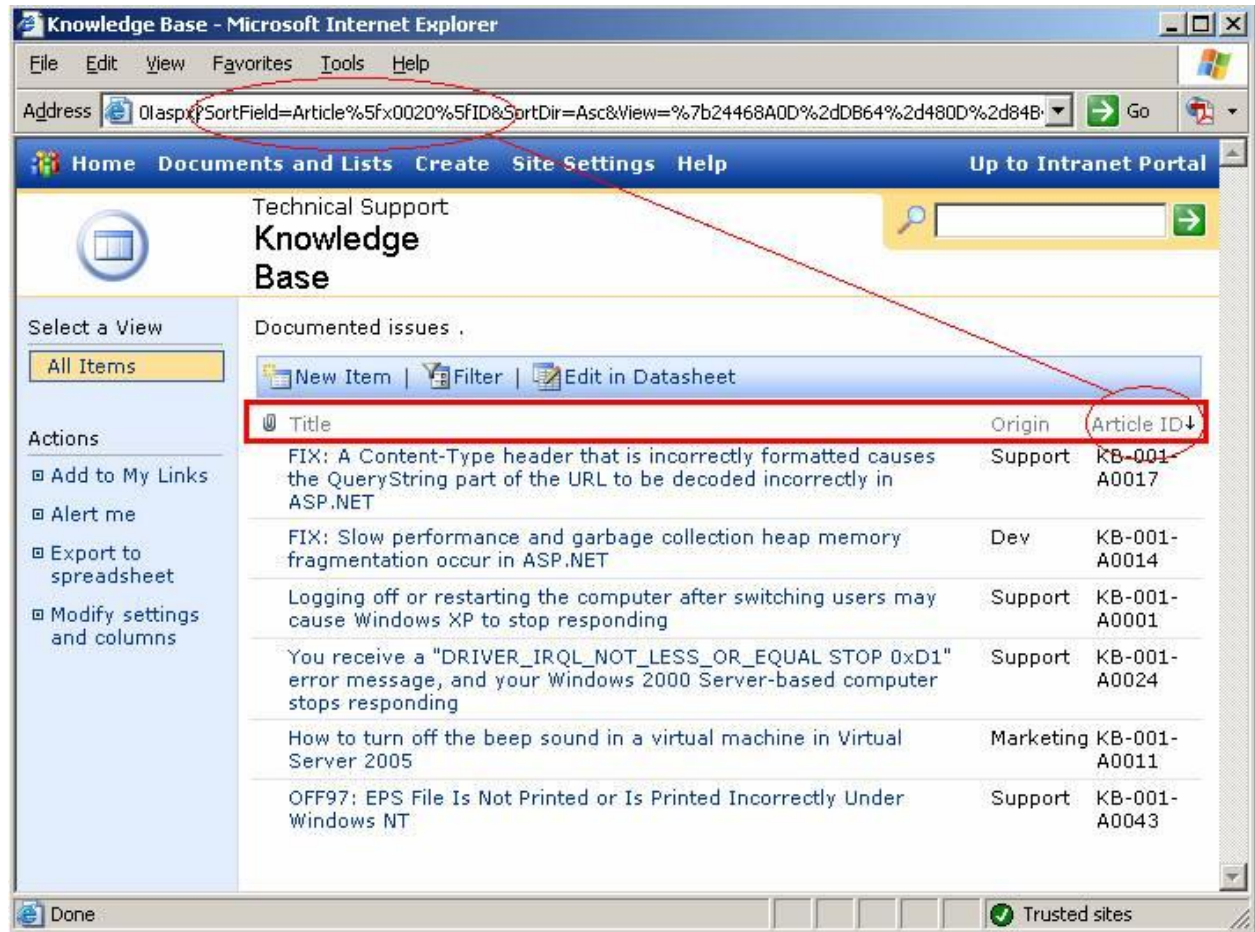
Metadata in Adobe PDF Files

It is not possible to view metadata information in *Adobe Acrobat Reader*, however several other PDF editors allow you to view the information. For example with *Adobe Acrobat*:



Metadata in SharePoint list items

In *SharePoint* lists, columns contain metadata.



In *SharePoint Document Libraries*, metadata about a document is stored in its *SharePoint* list item and in its internal structure (Word custom properties, PDF properties, etc.). Both are indexed but if a metadata name collision occurs, priority is given to the *SharePoint* item metadata.

The displayed name of the column may not be its real internal name, as needed by a Custom Field. To see the real name, sort the list by the column (click on the column's displayed name) and locate **SortField=Name** in the address bar of the browser. The name could be encoded with percent signs you must decode it. For example, a column's display name could be **My Column** but its real name is **My_x0020_Column**. In the address bar, it would be shown as **My%5fx0020%5fColumn**. The useful value to use as the custom field's metadata name is **My_x0020_Column**.

Custom Fields: Beware!

Here is the list of rules that must be followed when creating custom fields:

1. All custom fields that have the same name, even if they are in different sources, must be of the same type.
2. To change a custom field type:
 - a. Stop indexing all sources with the custom field.
 - b. Delete the custom field from all these sources.
 - c. Rebuild all these sources.
 - d. Compact the index.
 - e. Recreate the custom field for all these sources with the new data type.
 - f. Rebuild all these sources.