[MS-FSFIXML]:
FIXML Data Structure

Intellectual Property Rights Notice for Open Specifications Documentation

Revision Summary

Date	Revision History	Revision Class	Comments
11/06/2009	0.1	Major	Initial Availability
02/19/2010	1.0	Editorial	Revised and edited the technical content
03/31/2010	1.01	Editorial	Revised and edited the technical content
04/30/2010	1.02	Editorial	Revised and edited the technical content
06/07/2010	1.03	Editorial	Revised and edited the technical content
06/29/2010	1.04	Editorial	Changed language and formatting in the technical content.
07/23/2010	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
09/27/2010	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
11/15/2010	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
12/17/2010	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
03/18/2011	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
06/10/2011	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
01/20/2012	1.5	Minor	Clarified the meaning of the technical content.
04/11/2012	1.5	No change	No changes to the meaning, language, or formatting of the technical content.
07/16/2012	1.5	No change	No changes to the meaning, language, or formatting of the technical content.

Table of Contents

1 Introduction

This document specifies the FAST Index Markup Language (FIXML) data object format. This format is used to transmit and store documents submitted for indexing.

Sections 1.7 and 2 of this specification are normative and can contain the terms MAY, SHOULD, MUST, MUST NOT, and SHOULD NOT as defined in RFC 2119. All other sections and examples in this specification are informative.

1.1 Glossary

The following terms are defined in [MS-GLOS]:

UTF-8

The following terms are defined in [MS-OFCGLOS]:

anchor text
associated query
attribute vector
complete proximity boost
content collection
context catalog
crawled property
datetime
deep refinement
document identifier
document summary
document vector
FAST Index Markup Language (FIXML)
field importance level
full-text index context
full-text index field
hit highlighted summary
index schema
input summary class
internal property
item
item processing
managed property
occurrence boost
phrase break
property context
refiner
search clickthrough
search index
search service application
static rank
stemming by expansion
stemming by reduction
synthetic context catalog
tokenization
white space
word stacking

The following terms are specific to this document:

MAY, SHOULD, MUST, SHOULD NOT, MUST NOT: These terms (in all caps) are used as described in [RFC2119]. All statements of optional behavior use either MAY, SHOULD, or SHOULD NOT.

1.2   References

References to Microsoft Open Specifications documentation do not include a publishing year because links are to the latest version of the technical documents, which are updated frequently. References to other documents include a publishing year when one is available.

1.2.1   Normative References

We conduct frequent surveys of the normative references to assure their continued availability. If you have any issue with finding a normative reference, please contact dochelp@microsoft.com. We will assist you in finding the relevant information. Please check the archive site, http://msdn2.microsoft.com/en-us/library/E4BD6494-06AD-4aed-9823-445E921C9624, as an additional source.

[ISO-639] International Organization for Standardization, "Codes for the Representation of Names of Languages", ISO 639, http://www.loc.gov/standards/iso639-2/

[MS-FSCF] Microsoft Corporation, "Content Feeding Protocol Specification".

[MS-FSDP] Microsoft Corporation, "Document Processing Protocol Specification".

[MS-FSIN] Microsoft Corporation, "Input Normalization Data Structure".

[MS-FSSCFG] Microsoft Corporation, "Search Configuration File Format Specification".

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, http://www.rfc-editor.org/rfc/rfc2119.txt

1.2.2   Informative References

[MS-FSCX] Microsoft Corporation, "Configuration (XML-RPC) Protocol Specification".

[MS-FSIXDS] Microsoft Corporation, "Index Data Structures".

[MS-FSO] Microsoft Corporation, "FAST Search System Overview".

[MS-GLOS] Microsoft Corporation, "Windows Protocols Master Glossary".

[MS-OFCGLOS] Microsoft Corporation, "Microsoft Office Master Glossary".

1.3   Structure Overview (Synopsis)

This document specifies the format of the **FIXML** data structure. The data structure contains **items** prepared for indexing. One instance of the FIXML data structure represents the processed data of one item after **item processing**. The FIXML data structure represents the data payload of the protocol specified in [MS-FSDP].

The following figure provides a high-level overview of a search system that uses the FIXML data structure.
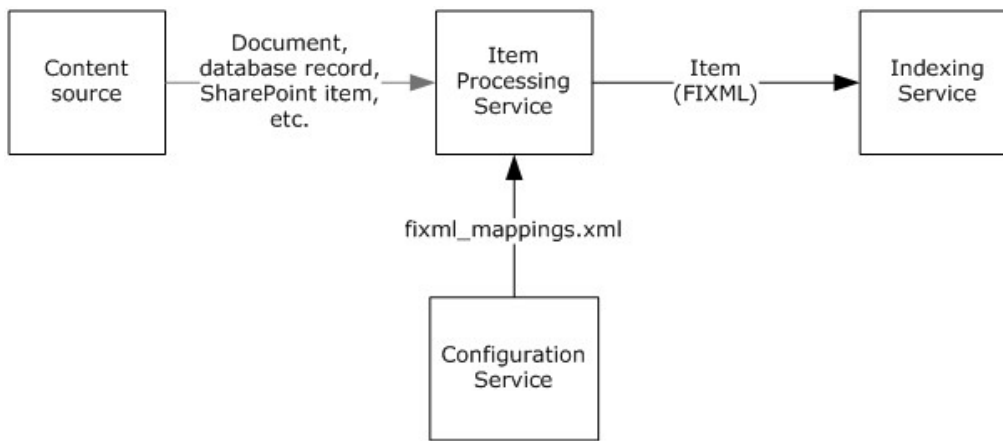
Figure 1: Overview

The search system first retrieves content from content sources such as Web servers, relational databases, file servers, or document repositories. An item processing service performs required processing to prepare the content for indexing. Examples of item processing are mapping of **crawled properties** to **managed properties**, parsing of document formats, linguistic normalization of text, and normalization of numeric data.

An [MS-FSDP] protocol client maps the processed managed properties into the appropriate FIXML data structures based on the mapping configuration, as described in the fixml_mappings.xml configuration file. The content of a managed property normally maps to more than one element in the FIXML data structure. For more information, refer to [MS-FSSCFG].

For more information about how the search system uses this data structure specification, refer to [MS-FSO].

The purpose of FIXML is to provide the data structure for a processed item and then to map the data to the **search index** structures that an **index schema** specifies. An entity that performs indexing consumes the FIXML data structure.

The FIXML data structure maps to the index data structure via the following index concepts:

- The **context catalog** represents a particular view of the searchable content. A numeric context catalog contains all numeric managed properties. A **synthetic context catalog** contains all text managed properties that can be queried individually. A **full-text index field** context catalog contains the index structures for a set of managed properties that accept ranked queries. Special-purpose context catalogs for **anchor text** and **associated query** data also exist. For more information about context catalog types, refer to section 2.5.2.1.

- The **property context** represents an individual managed property within the context catalog. The property context element in the FIXML contains the data of the managed property that will be indexed.

- The **document summary** elements in the FIXML are a textual representation of the managed property that can be returned with a query result.

- The **attribute vector** elements in the FIXML are used by the indexer to create data structures for **deep refinement** and full-text sorting.

1.4   Relationship to Protocols and Other Structures

The following figure provides a high-level overview of the services and protocols related to the FIXML data structure.
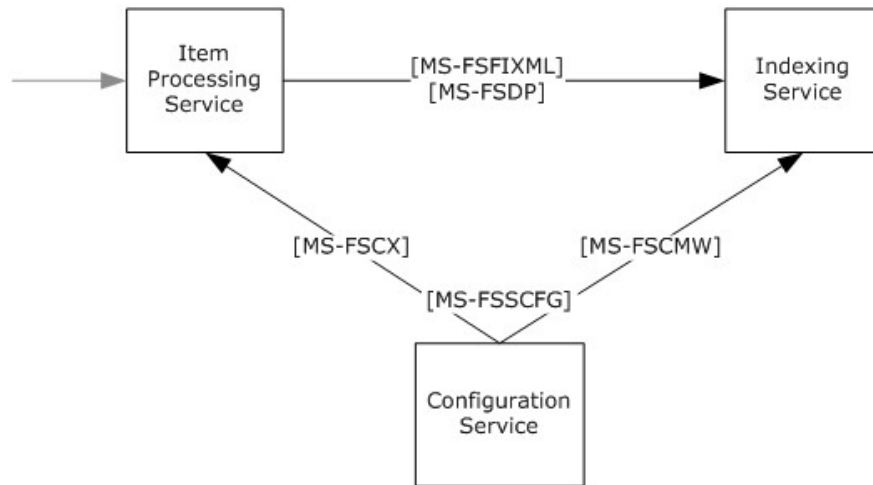
Figure 2: Service and protocol relationship

The FIXML data structure represents the data payload in [MS-FSDP]. A single data structure represents one item to be submitted for indexing. The indexing service uses the FIXML data structure to create the index data structures, as specified in [MS-FSIXDS] section 2.

The indexing service protocol client (described in [MS-FSDP]) creates the FIXML data structure by using the configuration information that is related to the index schema. The configuration information is in the fixml_mappings.xml configuration file. For more details, see [MS-FSSCFG].

The protocol client downloads the configuration file by using the protocol specified in [MS-FSCX].

1.5   Applicability Statement

The data structure represents items that are ready for indexing. The content of the data structure uses an XML document format and is represented as text. The document format applies for all full-text search applications that use the feeding protocol described in [MS-FSDP]. The feeding protocol creates index data structures.

1.6   Versioning and Localization

None.

1.7   Vendor-Extensible Fields

None.

2 Structures

2.1 Part Enumerations

None.

2.2 Extensions

None.

2.3 Global Elements

2.3.1 document

The **document** element represents one item to be submitted for indexing.

```
<xs:element name="document" type=" CT_document "/>
```

2.4 Global Attributes

None.

2.5 Complex Types

2.5.1 CT_document

Referenced by: **document**

A complex type that specifies one item to be submitted for indexing.

Child elements:

catalog: A **CT_catalog** element that specifies a context catalog. For a specification of context catalog types and required <catalog> elements, refer to section 2.5.2.1.

rank: A **CT_rank** element.

summary: A **CT_summary** element that specifies the **input summary class**.

attrVec: A **CT_attrVec** element that specifies a managed property formatted to facilitate deep refinement.

The following table describes the attribute for this type.

Name	Description
lang	The language code of the item, as specified by [ISO-639].

```
<xs:complexType name="CT_document">
  <xs:sequence>
    <xs:element name="catalog" type="CT_catalog" maxOccurs="unbounded"/>
    <xs:element name="rank" type="CT_rank"/>
    <xs:element name="attrVec" type="CT_attrVec" minOccurs="0"
```

Release: July 16, 2012

```
                                                 maxOccurs="unbounded"/>
        <xs:element name="summary" type="CT_summary"/>
      </xs:sequence>
      <xs:attribute name="lang" type="xs:string"/>
   </xs:complexType>
```

2.5.1.1 Item Requirements on Managed Properties

The following subsections specify item formatting requirements on managed properties that are mapped to the FIXML data structure.

The managed properties are mapped to one or more of the <catalog>, <summary>, and <attrVec> elements, depending on the index schema configuration. This section specifies requirements on the content of the managed properties before the mapping to the FIXML elements.

For a specification of how the managed properties are mapped to these elements, refer to the corresponding sections.

2.5.1.1.1 Document Vector

The "docvector" managed property contains the **document vector**. The property MUST contain zero or more name/value pairs. The name/value pairs are retrieved from the item during item processing.

The syntax of the "docvector" managed property is a string of successive name/value pairs that are enclosed in square brackets. The square brackets are a required part of the following syntax.

```
[<concept>, <weight>][<concept>, <weight>]…[<concept>, <weight>]
```

In the preceding syntax:

<concept> is a single word or a string that specifies a concept in the document vector.

<weight> is a floating point value between 0 and 1 that indicates the relevance of the concept.

The following syntax is an example of a "docvector" string.

```
[windows, 0.842][sharepoint, 0.736][microsoft office, 0.365148]
```

2.5.1.1.2 Document-Level Security Properties

The "docacl" and "docaclsystemid" managed properties MUST be formatted according to the following specification:

- If user access restrictions apply to the item, "docacl" and "docaclsystemid" managed properties MUST be formatted as specified in [MS-FSCF] section 2.2.28.

- If the item has no access restrictions (all users have read access), "docacl" and "docaclsystemid" managed properties MUST be set to the reserved word "all".

2.5.1.1.3 Anchor Text

The "anchortext" managed property represents the anchor text information for the item.

The managed property MUST contain zero or more anchor text strings. Each anchor text string represents the anchor text associated with one hyperlink that refers to this item. The managed property is mapped to the context catalogs named "bt1" and "anchortext", and as specified in a <map> setting in the fixml_mappings.xml configuration file. For a specification of this configuration file, refer to [MS-FSSCFG] section 2.10.

For details about mapping to the "anchortext" catalog, refer to section 2.5.2.1.4.

2.5.1.1.4 Associated Queries

The managed property named "assocqueries" represents the associated query information for the item.

The managed property MUST contain zero or more associated query strings. Each associated query string represents a user query string that resulted in a **search clickthrough** leading to this item. The managed property is mapped to the context catalogs named "bt1" and "assocqueries", and as specified in a <map> setting in the fixml_mappings.xml configuration file. For a specification of this configuration file, refer to [MS-FSSCFG] section 2.10.

For details about mapping to the context catalog named "assocqueries", refer to section 2.5.2.1.4.

2.5.1.1.5 Static Rank

Managed properties related to **static rank** SHOULD be set in the item processing as follows:

"docrank": The quality of the document based on link analysis.

"siterank": The quality of the document's corresponding site based on link analysis.

"urldepthrank": Inverse to the depth of the URL of the document.

If set, the value for each managed property MUST be an integer value that represents the rank contribution for this aspect of the static rank. The actual value range is implementation specific.

2.5.2 CT_catalog

Referenced by: **CT_document**

A complex type that specifies one context catalog in the search index. For further details, refer to [MS-FSSCFG] section 2.8.3.3.

Child element:

context: A **CT_context** element that specifies one property context within a context catalog. The <map> element in the fixml_mappings.xml configuration file identifies the <context> element to include. For details about the configuration file, see [MS-FSSCFG] section 2.10.

The **type** attribute of the <map> element specifies the mapping. When **type** is set to the value "context", the managed property content that this <map> element specifies MUST be mapped to a <context> element.

The following table describes the attribute for this type.

Name	Description
name	The name of the context catalog. This name MUST refer to a context catalog with the same name

Name	Description
	in the indexConfig.xml file, as specified in [MS-FSSCFG] section 2.10.3.3.

```
<xs:complexType name="CT_catalog">
  <xs:sequence>
    <xs:element name="context" type="CT_context" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="name" type="xs:string"/>
</xs:complexType>
```

2.5.2.1 Context Catalog Types

The context catalog names correspond to the value of the **dstcatalog** attribute in the fixml_mappings.xml <map> element when **type** is set to "context".

For a specification of context catalogs (<catalog> elements) that MUST be present for a particular index schema configuration, see [MS-FSSCFG] section 2.8.5.

For further details about context catalog types, see [MS-FSSCFG] section 2.10.3.2.

2.5.2.1.1 Synthetic Context Catalog (bt1)

A <catalog> element named "bt1" MUST be present in the FIXML. This represents a synthetic context catalog. The <context> entries inside this <catalog> element represent the non-numeric managed properties.

2.5.2.1.2 Numeric Context Catalog (bi1)

A <catalog> element named "bi1" MUST be present in the FIXML, with <context> element child objects representing the numeric managed properties.

Floating-point managed properties MUST use "." as the decimal separator.

Datetime managed properties MUST be encoded in the following format:

```
"[-][Y]YYYY-MM-DDThh:mm:ss[.d{1,7}]Z"
```

The datetime format legend is as follows:

- The letter *Y* indicates one cipher in the year (four or five digits accepted). The year specification can be negative. The minimum year is "-29000". The maximum year is "29000".

- The letter *M* indicates one cipher in the month number.

- The letter *D* indicates one cipher in the day number within the month.

- The letter *h* indicates one cipher in the hour number.

- The letter *m* indicates one cipher in the minute number.

- The letter *s* indicates one cipher in the second number. Seconds can optionally be specified with up to seven digits of decimal precision.

- The letters *T* and *Z* are literal constants.

2.5.2.1.3   Context Catalog for Item Metadata (meta)

A <catalog> element named "meta" MUST be present in the FIXML, with two <context> elements representing the following **internal properties**:

- **collection**: MUST contain the name of the **content collection** that this item is assigned to.

- **contentid**: MUST contain the **document identifier(3)** for the item.

The context catalog does not have any corresponding <map> entries in the fixml_mappings.xml configuration file.

2.5.2.1.4   Context Catalog for Anchor Text

The context catalog for anchor text contains a **complete proximity boost** representation of the "anchortext" managed property. For details about this managed property, refer to section 2.5.1.1.3.

The <context> elements within this <catalog> element MUST include the attribute **xml:lang**="complete".

This setting indicates that all **white space** characters in each anchor text string in the managed property MUST be replaced with the **UTF-8** character C2 BA (masculine ordinal indicator) in the <context> element. For details about white space characters, see [MS-FSIN] section 2.2.

Each anchor text string MUST be separated with the **phrase break** character represented by the UTF-8 character c7 82 (alveolar click) in the <context> element. This character instructs the indexer to treat each anchor text string as one word in the search index.

A <catalog> element named "anchortext" MUST be present. The <catalog> element MUST contain a <context> element named "canchortext" if any anchor text information is associated with the item. For details, see section 2.5.1.1.3.

A <catalog> element named "assocqueries" MUST be present. If any associated query information is associated with the item, the <catalog> element MUST contain a <context> element named "cassocqueries". For details, see section 2.5.1.1.4.

2.5.2.1.5   Context Catalog for Associated Queries

The context catalog for associated queries contains a complete proximity boost representation of the "assocqueries" managed property. For details about this managed property, refer to section 2.5.1.1.4.

The <context> elements within this <catalog> element MUST include the attribute **xml:lang**="complete".

This setting indicates that all white space characters in each associated query string in the managed property MUST be replaced with the UTF-8 character C2 BA (masculine ordinal indicator) in the <context> element. For details about white space characters, see [MS-FSIN] section 2.2.

Each associated query string MUST be separated with the phrase break character represented by the UTF-8 character c7 82 (alveolar click) in the <context> element. This character instructs the indexer to treat each associated query string as one word in the search index.

A <catalog> element named "assocqueries" MUST be present. If any associated query information is associated with the item, the <catalog> element MUST contain a <context> element named "cassocqueries". For details, see section 2.5.1.1.4.

2.5.2.1.6   Context Catalogs for Full-Text Index Field

The FIXML MUST contain one full-text index field <catalog> element for each full-text index field context catalog defined in the index schema. For further details about context catalog structure, see [MS-FSSCFG] section 2.8.5.

The <map> elements in fixml_mappings.xml that have a context catalog name (**dstcatalog** attribute) that starts with "bcat" specify managed property mapping to a full-text index field. For details about fixml_mappings.xml syntax, see [MS-FSSCFG] section 2.10.3.2.1.

The destination **full-text index context**, specified by the <map> attribute **dst**, is one of eight reserved property context names: "bconf1", "bconf2", "bconf3", "bconf4", "bconf5", "bconf6", "bconf7", and "bconf8". These eight property contexts are associated with the **field importance level** for the managed property. Multiple managed properties can be mapped to the same full-text index context. In this case, the content for each managed property MUST be included in a separate <context> element within the <catalog> element.

2.5.3   CT_context

Referenced by: **CT_catalog**

A complex type that specifies one property context within a context catalog. For further details, see [MS-FSSCFG] section 2.8.3.4.

Child elements:

sep: A **CT_sep** element that specifies an optional field separation within a <context> element.

a: A simple element that specifies grammatical variations for a word. For details, see section 2.5.3.1.3.

The following table describes the attributes for this type.

Name	Description
name	The name of the property context. The name MUST be equal to the value of a valid <map> attribute **dst** in fixml_mappings.xml. For further details about fixml_mappings.xml syntax, refer to [MS-FSSCFG] section 2.10.3.2.2. For details about mapping from managed properties to property contexts, refer to section 2.5.3.1.
numOccs	An optional occurrence count for the content of this <context> element. Setting the **numOccs** attribute to a value N for a context element that contains the word X has the same effect as repeating the word X N times inside the <context> element. This has a corresponding effect on the **occurrence boost** of the content inside the <context> element.
xml:lang	Specifies whether the consumer of the FIXML data structure will perform tokenization. If the attribute is not set, tokenization is already performed for the managed property as part of item processing. The usage of this attribute deviates from the definition in the XML namespace. The attribute is used for a specific purpose that is specified later. This is a deviation from XML schema rules, and hence this attribute is not specified in the XML schema. Valid values for the attribute are as follows: "space": The item content passed in the FIXML <context> element is not tokenized. For

Name	Description
	details, see section 2.5.3.1.1.
	"substring": The item content passed in the FIXML <context> element is not tokenized. For details, see section 2.5.3.1.1.
	"complete": **Tokenization** is already performed for the managed property as part of item processing. This setting MUST be used for the <context> elements named "canchortext" and "cassocqueries". For details, see sections 2.5.1.1.3 and section 2.5.1.1.4.
	If **xml:lang** is not present, this equals the setting **xml:lang**="space".

```
<xs:complexType name="CT_context" mixed="true">
  <xs:choice minOccurs="0" maxOccurs="unbounded">
    <xs:element name="sep" type="CT_sep"/>
    <xs:element name="a"/>
  </xs:choice>
  <xs:attribute name="name" type="xs:string" use="required"/>
  <xs:attribute name="numOccs">
    <xs:simpleType>
      <xs:restriction base="xs:unsignedInt">
        <xs:minInclusive value="1"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
</xs:complexType>
```

2.5.3.1   Content Requirements for Property Context

The mapping configuration specified in the following subsections specifies how the content mapping to the FIXML <catalog> element MUST be performed based on configuration data in fixml_mappings.xml and FieldProperties.xml.

2.5.3.1.1   Rules for Item Processing

The FieldProperties.xml configuration file is specified in [MS-FSSCFG] section 2.12 and contains item processing rules for managed properties.

The following table specifies the rules for item processing that MUST be applied on managed property data, before the mapping of the data to FIXML <context> elements. The **Parameter** column indicates the element or element/@attribute in FieldProperties.xml that specifies the configuration parameter. The **Item processing** column specifies the corresponding item processing that MUST be performed.

Parameter	Item processing
generic-tokenization	If this element is present, tokenization MUST NOT be applied to the managed property data before the mapping to the FIXML <context> element. For this tokenization mode, generic (language-independent) tokenization is performed inside the **search service application**.
	The consumer of the FIXML data structure MUST tokenize the content of the <context> element, as specified in [MS-FSIN] section 2.2.
	The <context> element MUST include the attribute **xml:lang**="space".
substring-tokenization	If this element is present, tokenization MUST NOT be applied to the managed property data before the mapping to the FIXML <context> element. For this tokenization mode, substring tokenization is performed

Parameter	Item processing
	inside the search service application. The consumer of the FIXML data structure MUST tokenize the content of the <context> element by using substring tokenization. The <substring-tokenization> attribute **N** specifies the n-gram value. The <context> element MUST include the attribute **xml:lang**="substring". For further details, see [MS-FSSCFG] section 2.8.3.5.
language-tokenization	If this element is present, tokenization MUST be applied to the managed property data before the mapping to the FIXML <context> element. For a specification of required tokenization, see [MS-FSIN] section 2.1. The <context> element MUST NOT include the attribute **xml:lang**. For the "url" and "urls" managed properties, tokenization MUST NOT be applied as part of item processing, even if the <language-tokenization> element is set for the properties in FieldProperties.xml. The consumer of the FIXML data structure MUST tokenize the content of the <context> element, as specified in [MS-FSIN] section 2.1.
language-tokenization/@lemmatization	If this element is present, stemming MUST be applied to the managed property data before the mapping to the FIXML <context> element. For a specification of required stemming, see section 2.5.3.1.3.
result/@type	If **type**="static" or **type**="dynamic", the <summary> element MUST include an <sField> element for this managed property that MUST be formatted as specified in section 2.5.7.1.

2.5.3.1.2   FIXML Mapping Rules for Managed Properties

The fixml_mappings.xml configuration file is specified in [MS-FSSCFG] section 2.10 and contains FIXML mapping rules for managed properties.

The following table specifies FIXML mapping rules based on particular values of the <map> attributes in fixml_mappings.xml.

<map> attribute	Item mapping rule
src	Specifies the name of the source managed property. For a specification of **src** naming rules, see [MS-FSSCFG] section 2.10.3.2.
maxsize	Managed properties larger than this size limit MUST be truncated before the mapping to the corresponding FIXML <context> element. The truncation MUST be applied to a word boundary.
phrasebreak	**phrasebreak**="yes": Semantically separated regions in the managed property text MUST be separated by the UTF-8 character c7 82 (alveolar click) in the <context> element. This phrase break character indicates a phrase break in the text flow. The query evaluation can use this character to avoid phrase match across such boundaries.
fieldseparationlength	If this attribute is present and contains a nonzero value, a <sep> element with **num**="*fieldseparationlength value*" MUST be added to the end of the FIXML <context> element. For details, see section 2.5.3.1.3.

2.5.3.1.3   Handling of Grammatical Variations

When a grammatical variation exists for a word, the <a> element MUST be included in the <context> element, replacing the word itself in the text flow. The <a> element indicates that all words inside the element are indexed through **word stacking**.

All words inside the <a> element MUST have a single uppercase suffix character added to the word:

- T: The original word as it appears in the managed property.

- L: Grammatical variations or the base form. The use depends on the stemming mode.

The L-suffix words added to the <a> element depend on the stemming mode specified in [MS-FSIN] section 2.3.

Stemming MUST be applied according to the language of the textual content of the managed property for each word (noun, adjective, verb, and pronoun) that has one or more grammatical variations.

Stemming form="variations" indicates **stemming by expansion**. All grammatical variations of the word MUST be added as L-suffix words, including the original word form as it appears in the managed property.

Stemming form="base form" indicates **stemming by reduction**. The canonical base form of the word MUST be added as an L-suffix word.

For a specification of the language-specific stemming mode to be applied, refer to [MS-FSIN] section 2.3. Stemming MUST be applied according to the language of the textual content of the managed property for each word (noun, adjective, verb, and pronoun) that has one or more grammatical variations.

2.5.4   CT_sep

Referenced by: **CT_context**

A complex type that specifies an optional field separation specification within a <context> element. The element instructs an indexer to insert **num** number of invisible words between the word before and after the <sep> element in the search index.

The following table describes the attribute for this type.

Name	Description
num	The number of invisible words to insert.

```
<xs:complexType name="CT_sep">
  <xs:attribute default="1" name="num" type="xs:unsignedShort"/>
</xs:complexType>
```

2.5.5   CT_rank

Referenced by: **CT_document**

A complex type that specifies a <rank> element with a fixed value, as specified in the XML schema. The consumer of the FIXML data structure MUST ignore this element.

```
<xs:complexType name="CT_rank">
  <xs:attribute name="class" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="dummy"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
</xs:complexType>
```

2.5.6 CT_summary

Referenced by: **CT_document**

A complex type that specifies a set of document summary fields, as specified in fixml_mappings.xml. For a specification of this configuration file, see [MS-FSSCFG] section 2.10.

Child element:

sField: A **CT_Sfield** element that specifies a document summary for the managed property. The <map> elements in the fixml_mappings.xml configuration file indicate which <sField> elements will be included. For a specification of this configuration file, see [MS-FSSCFG] section 2.10.

The **type** attribute of the <map> element specifies the mapping to <sField> elements. The managed property content that this <map> element specifies MUST be mapped to an <sField> element.

The following table describes the attribute for this type.

Name	Description
class	The name of the input summary class. It MUST be set to "content".

```
<xs:complexType name="CT_summary">
  <xs:sequence>
    <xs:element name="sField" type="CT_sField" minOccurs="1"
                                            maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="class" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="content"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
</xs:complexType>
```

2.5.7 CT_sField

Referenced by: **CT_summary**

A complex type that specifies a document summary for the managed property, as specified by the **name** attribute.

The following table describes the attribute for this type.

Name	Description
name	The name of the document summary. It MUST be equal to the value of the **dst** attribute to the corresponding <map> element of fixml_mappings.xml.

```
<xs:complexType name="CT_sField">
  <xs:attribute name="name" type="xs:string" use="required"/>
</xs:complexType>
```

2.5.7.1   Mapping Rules for Document Summary

The <map> elements in the fixml_mappings.xml configuration file that have property context names that start with "bsum" or "bsrc" specify a managed property mapping to an <sField> element. For a specification of this configuration file, refer to [MS-FSSCFG] section 2.10. For details about the naming of index field prefixes, see [MS-FSSCFG] section 2.1.2.

The document summary <sField> elements that have names that start with "bsrc" represent input to creation of a **hit highlighted summary**. The <sfield> element MUST include the following additional formatting markup:

▪ The UTF-8 character c7 82 (alveolar click) MUST separate semantically separated regions in the text. This character indicates a phrase break in the text flow. The hit highlighted summary feature can use it to avoid phrase match across such boundaries.

2.5.8   CT_attrVec

Referenced by: **CT_document**

A complex type that specifies a set of attribute vector data (<avField> elements).

Child element:

avField: An **xs:string** element that specifies one attribute value. If no attribute value exists for the managed property in this item, an empty <avField> element MUST be present.

The <map> elements in the fixml_mappings.xml configuration file identifies the <attrVec> elements to include. For a specification of this configuration file, see [MS-FSSCFG] section 2.10.

The **type** attribute of the <map> element specifies the mapping. When **type** is set to "attributevector", the managed property content that this <map> element specifies MUST be mapped to an <attrVec> element.

The following table describes the attribute for this type.

Name	Description
name	The name of the attribute vector. It MUST be equal to the value of the **dst** attribute to the corresponding <map> element of fixml_mappings.xml. For a specification of attribute vector types and associated naming rules, refer to [MS-FSSCFG] section 2.10.3.2.1. **dst**="batv*managedPropertyName*" for a managed property named *managedPropertyName* that have full-text sort enabled. **dst**="bavn*managedPropertyName*" for a managed property *managedPropertyName* that have an associated **refiner**.

```
<xs:complexType name="CT_attrVec">
  <xs:sequence>
```

```
      <xs:element maxOccurs="unbounded" name="avField" type="xs:string"/>
    </xs:sequence>
    <xs:attribute name="name" type="xs:string" use="required"/>
  </xs:complexType>
```

2.5.8.1   Mapping Rules for Attribute Vector

The following table specifies FIXML mapping rules based on particular values of the <map> attributes in fixml_mappings.xml. For a specification of this configuration file, see [MS-FSSCFG] section 2.10.

<map> attribute	Item mapping rule
src	Specifies the name of the source managed property. For a specification of **src** naming rules, see [MS-FSSCFG] section 2.10.3.2.
multi	If this attribute is set to "yes", the source managed property supports multiple (text or numeric) values for attribute vectors. Each value MUST be inserted as individual **attrVec**/**avField** elements in the FIXML.
defaultvalue	If this attribute is set and not empty, and the source managed property is empty, the associated **attrVec/avField** element MUST be set to the value of **defaultvalue**.

2.6   Simple Types

None.

3   Structure Examples

3.1   FIXML Structure

The following example illustrates one FIXML <document> element that represents one item that is sent for indexing.

Content for indexing in the full-text index field named "content". In this example, "body" is mapped to "bconf1", and "urls" is mapped to "bconf4". Note the special handling of "urls", as specified in section 2.5.3.1.1.

The following examples indicate the phrase break UTF-8 character c7 82 (alveolar click) as "ǂ ".

Stemming is shown via the <a> element.

```
<document lang="en">
  <catalog name="bcatcontent">
    <context name="bconf1"> ǂ <a>welcomeT welcomeL welcomesL</a> to the Official contoso
<a>pharmaceuticalsT pharmaceuticalL pharmaceuticalsL</a> <a>websiteT websiteL websitesL</a>
the best <a>choiceT choiceL choices</a> in <a>pharmaceuticalsT pharmaceuticalL
pharmaceuticalsL</a> ǂ <sep num="256"/></context>
    <context name="bconf4">
      <![CDATA[ ǂ http://www.contoso.com/welcome.html ǂ ]]><sep num="256"/></context>
    <context name="bconf7"> ǂ <a>contosoT contosoL</a> ǂ
            <sep num="256"/></context>
  </catalog>
```

The following example shows content for indexing in the synthetic context catalog named "bt1". This contains all non-numeric managed properties that are configured as individually searchable.

In this example, "docacl" and "docaclsystemid" are set to "all". This indicates that no restrictions on access rights exist on the item.

```
<catalog name="bt1">
  <context name="bcontitle"> ǂ <a>contosoT contosoL</a> ǂ </context>
  <context name="bconanchortext"><![CDATA[]]></context>
  <context name="bconassocqueries"><![CDATA[]]></context>
  <context name="bconcontenttype"><![CDATA[ ǂ text/html ǂ ]]></context>
  <context name="bconformat"><![CDATA[ ǂ Web Page ǂ ]]></context>
  <context name="bconlanguage"><![CDATA[ ǂ en ǂ ]]></context>
  <context name="bconlanguages"><![CDATA[ ǂ en ǂ ]]></context>
  <context name="bconcharset"><![CDATA[ ǂ utf-8 ǂ ]]></context>
  <context name="bconurls">
     <![CDATA[ ǂ http://www.contoso.com/welcome.html ǂ ]]></context>
  <context name="bcondocacl"><![CDATA[ ǂ all ǂ ]]></context>
  <context name="bcondocaclsystemid"><![CDATA[ ǂ all ǂ ]]></context>
  <context name="bconcrawledpropertiescontent"></context>
  <context name="bconcrawledpropertynames"><![CDATA[]]></context>
  <context name="bconcompanies">
     <![CDATA[ ǂ Fabrikam ǂ Woodgrove Bank ǂ Contoso Pharmaceuticals ǂ ]]></context>
  <context name="bconlocations"><![CDATA[ ǂ Contosoville ǂ ]]></context>
  <context name="bconpersonnames"><![CDATA[]]></context>
  <context name="bconemails"><![CDATA[]]></context>
  <context name="bcondates"><![CDATA[ ǂ 2004-04-29 ǂ ]]></context>
  <context name="bcontimes"><![CDATA[]]></context>
  <context name="bconprices"><![CDATA[]]></context>
```

```
    <context name="bconextractedurls">
        <![CDATA[ ǂ www.contoso.com ǂ  www.fabrikam.com ǂ ]]></context>
</catalog>
```

The special context catalog named "assocqueries" contains query terms that have led to a clickthrough to this item from a query hit list, as shown in the following example.

```
<catalog name="assocqueries">
    <context name="cassocqueries" xml:lang="complete" numOccs="35">
        <![CDATA[contoso]]></context>
</catalog>
```

The special context catalog named "anchortext" contains anchor texts that reference this item, as shown in the following example.

```
<catalog name="anchortext">
    <context name="canchortext" xml:lang="complete" numOccs="2">
        <![CDATA[contoso]]></context>
</catalog>
```

The following example shows content for indexing in the numeric context catalog named "bi1". This contains all numeric managed properties.

```
<catalog name="bi1">
    <context name="bconprocessingtime">
        <![CDATA[2009-02-16T16:44:02Z]]></context>
    <context name="bcondocdatetime"><![CDATA[2004-10-29T15:19:23Z]]></context>
    <context name="bconsize"><![CDATA[28601]]></context>
    <context name="bcondocrank"><![CDATA[0]]></context>
    <context name="bconsiterank"><![CDATA[0]]></context>
    <context name="bconurldepthrank"><![CDATA[630]]></context>
</catalog>
```

The following example shows item metadata ("contentid" and "collection").

```
<catalog name="meta">
    <context name="contentid">
        <![CDATA[ ǂ http://www.contoso.com/welcome.html ǂ ]]><sep num="256"/></context>
    <context name="collection" xml:lang="space">
        <![CDATA[test1]]><sep num="256"/></context>
</catalog>
```

The following example shows the dummy element that MUST be present.

```
<rank class="dummy">
</rank>
```

The following example shows attribute vectors for managed properties that have full-text sorting enabled.

```
<attrVec name="batvtitle">
```

```
    <avField><![CDATA[Contoso Pharmaceuticals: www.contoso.com]]></avField>
  </attrVec>
  <attrVec name="batvprocessingtime">
    <avField><![CDATA[2009-02-16T16:44:02Z]]></avField>
  </attrVec>
  <attrVec name="batvdocdatetime">
    <avField><![CDATA[2004-10-29T15:19:23Z]]></avField>
  </attrVec>
  <attrVec name="batvsize">
    <avField><![CDATA[28601]]></avField>
  </attrVec>
  <attrVec name="batvdocrank">
    <avField><![CDATA[0]]></avField>
  </attrVec>
  <attrVec name="batvsiterank">
    <avField><![CDATA[0]]></avField>
  </attrVec>
  <attrVec name="batvurldepthrank">
    <avField><![CDATA[630]]></avField>
  </attrVec>
```

The following example shows attribute vectors for managed properties that have an associated refiner.

```
  <attrVec name="bavnlanguages">
    <avField><![CDATA[en]]></avField>
  </attrVec>
  <attrVec name="bavncompanies">
    <avField><![CDATA[Fabrikam]]></avField>
    <avField><![CDATA[Woodgrove Bank]]></avField>
    <avField><![CDATA[Contoso Pharmaceuticals]]></avField>
  </attrVec>
  <attrVec name="bavnlocations">
    <avField><![CDATA[Contosoville]]></avField>
  </attrVec>
  <attrVec name="bavnpersonnames">
    <avField><![CDATA[]]></avField>
  </attrVec>
  <attrVec name="bavnemails">
    <avField><![CDATA[]]></avField>
  </attrVec>
  <attrVec name="bavndates">
    <avField><![CDATA[2004-04-29]]></avField>
  </attrVec>
  <attrVec name="bavntimes">
    <avField><![CDATA[]]></avField>
  </attrVec>
  <attrVec name="bavnextractedurls">
    <avField><![CDATA[www.contoso.com]]></avField>
    <avField><![CDATA[www.fabrikam.com]]></avField>
  </attrVec>
  <attrVec name="bavnprices">
    <avField><![CDATA[]]></avField>
  </attrVec>
  <attrVec name="bavnformat">
    <avField><![CDATA[Web Page]]></avField>
  </attrVec>
```

```
    <attrVec name="bavncrawledpropertynames">
    </attrVec>
    <attrVec name="bavnsize">
      <avField><![CDATA[28601]]></avField>
    </attrVec>
    <attrVec name="bavndocdatetime">
      <avField><![CDATA[2004-10-29T15:19:23Z]]></avField>
    </attrVec>
```

The following example shows content for document summaries.

```
    <summary class="content">
      <sField name="internalid">
        <![CDATA[7894d0895b21e07148fbcfe1471d83ca_test1]]></sField>
      <sField name="contentid">
        <![CDATA[http://www.contoso.com/welcome.html]]></sField>
      <sField name="collection"><![CDATA[test1]]></sField>
      <sField name="contentids">
        <![CDATA[http://www.contoso.com/welcome.html]]></sField>
      <sField name="bsrctitle">
        <![CDATA[Contoso Pharmaceuticals: www . contoso . com]]></sField>
      <sField name="bsrcbody">
        <![CDATA[Welcome to the Official Contoso Pharmaceuticals Website. The best choice in
 pharmaceuticals.]]></sField>
      <sField name="bsumteaser">
        <![CDATA[The best choice in pharmaceuticals]]></sField>
      <sField name="bsumcontenttype"><![CDATA[text/html]]></sField>
      <sField name="bsumformat"><![CDATA[Web Page]]></sField>
      <sField name="bsumlanguage"><![CDATA[en]]></sField>
      <sField name="bsumlanguages"><![CDATA[en]]></sField>
      <sField name="bsumcharset"><![CDATA[utf-8]]></sField>
      <sField name="bsumurls">
        <![CDATA[http://www.contoso.com/welcome.html]]></sField>
      <sField name="bsumurl">
        <![CDATA[http://www.contoso.com/welcome.html]]></sField>
      <sField name="bsumprocessingtime"><![CDATA[2009-02-16T16:44:02Z]]></sField>
      <sField name="bsumdocdatetime"><![CDATA[2004-10-29T15:19:23Z]]></sField>
      <sField name="bsumsize">28601</sField>
      <sField name="bsumdocrank">0</sField>
      <sField name="bsumsiterank">0</sField>
      <sField name="bsumurldepthrank">630</sField>
      <sField name="bsumdocacl"><![CDATA[all]]></sField>
      <sField name="bsumdocaclsystemid"><![CDATA[all]]></sField>
      <sField name="bsumcrawledpropertynames"><![CDATA[]]></sField>
      <sField name="bsumcompanies">
        <![CDATA[Fabrikam; Woodgrove Bank; Contoso Pharmaceuticals]]></sField>
      <sField name="bsumlocations"><![CDATA[Carlsberg]]></sField>
      <sField name="bsumpersonnames"><![CDATA[]]></sField>
      <sField name="bsumemails"><![CDATA[]]></sField>
      <sField name="bsumdates"><![CDATA[2004-04-29]]></sField>
      <sField name="bsumtimes"><![CDATA[]]></sField>
      <sField name="bsumprices"><![CDATA[]]></sField>
      <sField name="bsumextractedurls">
        <![CDATA[www.contoso.com; www.fabrikam.com]]></sField>
    </summary>
  </document>
```

4 Security Considerations

None.

5   Appendix A: Full XML Schemas

For ease of implementation, this section provides the full World Wide Web Consortium (W3C) XML schemas for the elements, attributes, complex types, and simple types described in the preceding sections.

```xml
<?xml version="1.0" encoding="utf-8" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

  <!-- ********************** Global elements ********************** -->

  <xs:element name="document" type="CT_document"/>

  <!-- ********************** Complex types ********************** -->

  <xs:complexType name="CT_document">
    <xs:sequence>
      <xs:element name="catalog" type="CT_catalog" maxOccurs="unbounded"/>
      <xs:element name="rank" type="CT_rank"/>
      <xs:element name="attrVec" type="CT_attrVec" minOccurs="0"
                                                   maxOccurs="unbounded"/>
      <xs:element name="summary" type="CT_summary"/>
    </xs:sequence>
    <xs:attribute name="lang" type="xs:string"/>
  </xs:complexType>

  <xs:complexType name="CT_catalog">
    <xs:sequence>
      <xs:element name="context" type="CT_context" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="name" type="xs:string"/>
  </xs:complexType>

  <xs:complexType name="CT_context" mixed="true">
    <xs:choice minOccurs="0" maxOccurs="unbounded">
      <xs:element name="sep" type="CT_sep"/>
      <xs:element name="a"/>
    </xs:choice>
    <xs:attribute name="name" type="xs:string" use="required"/>
    <xs:attribute name="numOccs">
      <xs:simpleType>
        <xs:restriction base="xs:unsignedInt">
          <xs:minInclusive value="1"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:complexType>

  <!-- Note that the xml:lang attribute is used for the context element but -->
  <!-- is not used according to the standard XML 1.0 namespace definition. -->

  <xs:complexType name="CT_sep">
    <xs:attribute default="1" name="num" type="xs:unsignedShort"/>
  </xs:complexType>

  <xs:complexType name="CT_rank">
    <xs:attribute name="class" use="required">
      <xs:simpleType>
```

```
          <xs:restriction base="xs:string">
            <xs:enumeration value="dummy"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
  </xs:complexType>

  <xs:complexType name="CT_summary">
    <xs:sequence>
      <xs:element name="sField" type="CT_sField" minOccurs="1"
                                            maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="class" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="content"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:complexType>

  <xs:complexType name="CT_sField">
    <xs:attribute name="name" type="xs:string" use="required"/>
  </xs:complexType>

  <xs:complexType name="CT_attrVec">
    <xs:sequence>
      <xs:element maxOccurs="unbounded" name="avField" type="xs:string"/>
    </xs:sequence>
    <xs:attribute name="name" type="xs:string" use="required"/>
  </xs:complexType>

</xs:schema>
```

6 Appendix B: Product Behavior

The information in this specification is applicable to the following Microsoft products or supplemental software. References to product versions include released service packs:

- Microsoft® FAST™ Search Server 2010

Exceptions, if any, are noted below. If a service pack or Quick Fix Engineering (QFE) number appears with the product version, behavior changed in that service pack or QFE. The new behavior also applies to subsequent service packs of the product unless otherwise specified. If a product edition appears with the product version, behavior is different in that product edition.

Unless otherwise specified, any statement of optional behavior in this specification that is prescribed using the terms SHOULD or SHOULD NOT implies product behavior in accordance with the SHOULD or SHOULD NOT prescription. Unless otherwise specified, the term MAY implies that the product does not follow the prescription.

7 Change Tracking

No table of changes is available. The document is either new or has had no changes since its last release.

8 Index