BIC Bites

@BIC1UK
www.bic.org.uk
info@bic.org.uk

BIC Bites are a series of short, high level, papers on topics of interest to BIC Members. If there is a topic you would like to see as a BIC Bite, please contact info@bic.org.uk

Choosing Keywords for Discoverability

This BIC Bite is relevant to metadata managers, and marketers responsible for optimising product discovery.

Introduction

For book buyers and readers, knowing what a book is *about* is one of the key factors in a purchasing decision. Most publishers provide information on the subject of their books to retailers as one part of a larger set of metadata, often in a standard format such as ONIX. The subject information is usually provided using a structured subject coding scheme such as the BIC subject codes or, more recently, *Thema*. But structured subject schemes are only one side of the coin, and providing *unstructured keywords* can greatly improve a book's online discoverability.

What are keywords?

Keywords are natural language words or short phrases that relate to the content or theme of a book, and which are used to supplement the structured metadata such as title, author, subject, an audience coding, and any descriptive text. They are words or phrases that are expected to be used as search terms by consumers searching for a book – particularly when the consumer does not know the exact title or author name, or when the consumer does not have a specific book in mind. Including a good list of relevant keywords in your book metadata ensures your book is more easily or more frequently discovered by potential purchasers within an online bookstore. It will appear higher up the list of search results, and will be shown more often to purchasers who are searching for that particular type of book.

What types of keywords should I choose?

- **Keywords don't replace structured subject information schemes** like BIC or *Thema*. They *supplement* it. Use keywords to provide *additional* highly detailed information that cannot be encoded with a structured subject code.
- Choose keywords carefully. **Use keywords that are search terms likely to be used by consumers and potential purchasers** looking for books like yours, so for example for fiction, the names of key characters, locations or settings, plot themes and genre terms, and for non-fiction, important 'terms of art' or jargon all make good keywords.
- Other data fields are often used as an additional source of keywords. Some retailers 'mine' long descriptions, tables of contents and so on for keywords. Repetition is unlikely to boost search rankings further, but you cannot guarantee that every retailer will do this text mining, so **do include important keywords even if you have already included them in descriptive text or in the table of contents**
- but **don't repeat words or phrases that are already present in one of the more important structured metadata fields**. Repeating the name of the author or a word from the title as a keyword is pointless. If a consumer happens to search on that word, they will find the book based on that word occurring in the title or contributor name fields – not because it is included as a keyword
- **Five well-chosen keywords are more effective than fifty poor selections, but there is no 'right' number**. Choose as many as you need to cover all the likely search terms. But if your list exceeds 20 or so terms, ensure each term is truly worthwhile. However, there is a practical limit to the number of keywords that can be carried within ONIX.
- **Highly specific keywords** work better than overly broad or generic terms.
- **Don't game the system**. Adding '50 shades', 'adult colouring book' or the name of a better-known author as keywords might seem like a good way of getting your book found by readers searching for other books. In reality, it is a good way of getting all your keywords ignored by retailers. Most retailers also have rules against using misleading or irrelevant phrases, and against including words like 'bestseller', 'free', 'newly published'.
- **There is no particular need to include plurals of nouns, closely-related words or misspellings**. Retailer search engines usually do some fuzzy matching to deal with misspellings, and they often use 'stemming', so that a search for 'fishing' or 'cycling' would match keywords like 'fish', 'fisherman' or 'bicycle'
- but **do include genuine alternative spellings, synonyms, acronyms and alternative or related phrases** – zucchini *and* courgette, eggplant *and* aubergine, diet, nutrition, healthy eating *and* weight loss, for a book of low-calorie vegetable recipes. A thesaurus can help.

- It can be useful to **reconsider and revise the list of keywords throughout a book's life**. New terminology or newsworthy topics become relevant and could be added to the keyword list.
- **Very rarely, some retailers might *display* the list of keywords provided by the publisher**, for example in a 'tag cloud' or as part of a facetted search; this is not at all common, but bear it in mind as you select keywords. Publishers and retailers might also add the same keywords to web pages, for SEO purposes.

How are keywords carried in ONIX?

Keywords in ONIX are carried as a simple, semi-colon separated list of words and phrases, within the <Subject> composite, using subject scheme identifier 20. For a book about the Arts and Crafts movement in English architecture, a set of keywords might look like this:

```
<Subject>
    <SubjectSchemeIdentifier>20</SubjectSchemeIdentifier>
    <SubjectHeadingText>Augustus Pugin; William Morris; Kelmscott House; John Ruskin;
    Philip Webb; Red House; Charles Voysey; Edwin Lutyens; Edward Schroeder Prior;
    Voewood; Euston fire station; vernacular architecture; medieval revival; gothic
    revival</SubjectHeadingText>
</Subject>
```

There is no difference between ONIX 2.1 and 3.0 [1] [2]. Note that use of the <MainSubject> composite in version 2.1, or the <MainSubject/> flag in 3.0 is not appropriate for keywords. Some data recipients will use only the first few keywords, so put your 'best' keywords first. Of course, the keywords merely supplement the appropriate range of BIC or *Thema* subject codes and qualifiers in separate repeats of the <Subject> and <MainSubject> composites. Note also that if the book were about any of the architects in particular (rather than an overview of the entire movement) then using <NameAsSubject> would be more appropriate.

Are there tools to help choose keywords?

The best source of keywords is undoubtedly editorial knowledge of the subject matter and the text of the book itself. The second best source of inspiration for keywords is 'reader reviews' in online bookstores. These show how real readers describe books. However, there are a few useful online tools to add to this:

- **Übersuggest** uses Google's search autocomplete system (also used by most web browsers) to suggest extended or slightly modified search phrases. The 'word cloud' option can provide inspiration for useful keywords. However, bear in mind that it's based on common Google searches, rather than searches where finding a book is the aim.
- **uClassify** allows you to check text extracts from a book to see how it can be classified – pasting in a few key paragraphs from the book can suggest keywords or classification headings to add. uClassify scans thousands of pre-classified documents to 'train' a computer system, then new text can be classified according to its similarity to various original training documents.
- **LibraryThing** shows tags associated with books by readers. Some of these simply duplicate the structured metadata, but others can be useful.

Above all, use these tools as a source of *ideas*, not a source of keywords, and 'think like your target customer'.

Written by Graham Bell, Executive Director, EDItEUR
Last updated 22nd September 2016

[1] ONIX 2.1 suggests a maximum length of the text in <SubjectHeadingText> as 100 characters, but in practice, exceeding this will not generally cause problems. ONIX 3.0 extends this to a more useful 500 characters, and exceeding 500 characters is not recommended.
[2] spaces following each semi-colon are recommended but not vital, and should not be relied upon by recipients.