

Introduction

Hyperlinks have an important role in indexes. A linked index to an online database of journals or bulletins may use hyperlinks. Back of book style indexes may use hyperlinks and/or sub addresses to make the index active. As more material is placed on the Web, it is being indexed, normally by search engines, now supported by increasingly effective Artificial Intelligence (AI). There is still a need for carefully curated indexes for high value journals and society archives. A survey has analysed the support provided by a variety of applications and browsers and discovered a wide variety of behaviours. Survey results are included. A set of requirements is proposed, which, if accepted, will give improved performance for a wide range of users. Cooperation is sought from publishers and software developers to make this happen.

History of the Web and Hyperlinks

Hypertext was used before computers became available as a mechanism for directing readers from one page to another. Users of motor car manuals experience hypertext when they are directed to a new page giving more information about a feature. The online versions are hyperlinked. Wikipedia has a page about hypertext fiction (Wikipedia 2023).

A history of the Web and hyperlinks can be found at (Hoffman, The History of the Web 2023) An early web database was an index of films IMDb, which still exists and is now owned by Amazon (IMDb 2023). There are complex legal issues involved with linking to copyrighted or commercially valuable material with some cases lasting for years (Hoffman, The Right to Link 2021).

(Jehoram 2016) references a judgement of the European Court of Justice Sanoma/GS Media (8 Sept 2016) about a posting by GS of a hyperlink to an as-yet unpublished edition of Playboy. The legal status was considered ambiguous with commercial organisations more likely to be committing an offence when there is a financial gain.

Definitions

A few definitions are introduced to assist the story.

Index

- a. A curated, human readable, ordered list of entries including locators enabling a user to find content in a work or collection of works. An example is a back of book index or an index to news bulletins or newspaper articles.
- b. A computer generated, ordered list of entries including pointers enabling fast location of content by an automated system. Examples are indexes created to speed access by relational databases, internet search engines, PDF readers, or disk drives.

Example a is used here although for a linked index, type b indexes will be operating in the background.

Linking

Linking allows a user to move easily from a location in a document to another location in the same or another document. A good linking strategy allows the user to return easily. Links are not hard wired like wiring in a city, they are created as needed usually by a search mechanism often aided by complex databases and lots of computing power.

Internal Linking

Internal linking allows a user to move between locations within a document. Some document systems including Word and PDF can use hyperlink sub addresses in their internal linking mechanism. Word allows internal linking to numbered items, headings, bookmarks, footnotes, endnotes, equations, figures, and tables. When exported to PDF the links are supported.

Many features can be used as link targets, but special processing is required if standard approaches are not used. The best targets are those with a fixed identity such as headings, captions, bookmarks, or shapes in Word. Page numbers and fields can be used when the document is stable.

Hyperlinks

Hyperlinks are a computerised implementation of hypertext and were proposed by Tim Berners-Lee, a British physicist then working at CERN, as the mechanism for transferring scientific information between collaborators. This was the beginning of the World Wide Web as we now know it. One of the first web addresses was <http://info.cern.ch/hypertext/WWW/TheProject.html> which still exists.

Display attributes

The text to display when a hyperlink is created is typically user defined. Long hyperlink addresses can be hidden behind convenient identifiers. Other attributes may be available including a screen tip. For indexes the text displayed will be a locator or part of a locator.

Deep Hyperlinks

When a hyperlink is to a website, a deep hyperlink is to a location within the website.

Fragments

The W3C specification defines a fragment identifier following a # symbol but leaves software developers to decide how they are used (W3C 2023). Some application, including Microsoft Word, call them sub addresses which can take users to specific pages or bookmarks within a document.

Sub-Addresses

A sub-address is an application of a fragment, defining a location within a document. For a document with page numbers an example is #page=100. For a document containing named ranges an example is #RangeName.

Sub addresses are not supported for all users, it depends on the hardware and software.

Page number sub address

#page=100 indicates a page number sub address. For a PDF file it will take the user to the 100th page. If

- a. the file contains frontmatter or;
- b. is divided into several sections, each beginning at page 1 or;
- c. is one of many files making up a book;

then the PDF page will not correspond to the page number the reader will recognise. An adjustment is required.

Range Name Sub Addresses

A range name sub address will take the user to the range. Depending on the applications involved it could be the beginning of the named range, the entire range, or a close location.

Embedding

Word processing and publishing software manage text layout and many other features using embedded codes. Fonts, styles, fields, bookmarks, tables, pictures, and shapes use embedded codes. A variety of index embedding methods are used, depending on the composition software (Ream 2016).

Fields

Fields are active or passive embedded codes. Some contain user data; some are updated as document content changes. Microsoft Word uses over 70 field types for page numbering, content lists, indexes, and other purposes.

Bookmarks

Bookmarks are used to by Microsoft Word to name ranges. Range names must be unique within the relevant scope, which could be several files if they are eventually to be merged.

Bookmarks are used differently by PDF files. Range names imported from Microsoft Word are respected, can be referenced from an index and via hyperlinks from other applications, depending on hardware and software, but cannot be readily viewed.

Hyperlinked Indexes

When using a hyperlinked index, you may be confronted with a guidance note like the one below.

The journals are in PDF format. You will need to have a suitable PDF reader installed to read the articles (e.g., Adobe Reader or Foxit Reader).
Depending on the browser you are using, the pdf reader you are using and the way your browser is configured to handle pdf files, clicking on the links in the following index can produce several different types of behaviour, such as:
The relevant journal will load at the page containing the desired article, or...
The relevant journal will load at the first page - you will then need to scroll to the required page, or...
The browser will present a dialogue box asking if you wish to 'Open' or 'Save' the journal. If you select 'Open', the journal will open at either the first page or the page containing the desired article. If you select 'Save', the journal will be downloaded to your 'downloads' folder and you will need to open it manually from that location, or...
The browser will download the journal to your 'downloads' folder without asking and you will need to open it manually from that location.

If the indexed documents are behind a pay wall or another protection mechanism, then it becomes even more complicated.

Test Document

The test document was the IndexExploit user guide. It contains three indexes, all generated from the embedded index entries. The first is a conventional generated index. The second is a synchronised index generated from a range within the document, the third is a linked index. IndexExploit is a Word add in that performs many functions to support linked indexes (IndexBase 2023). It's been used to create complex embedded indexes and to support large linked indexes. It was used to create the linked index that enabled the survey.

The Word document was exported to PDF using Adobe Acrobat Pro.

Behaviour of a Linked Index using Microsoft Word 365

The embedded index entry used in the survey is shown below.

XE "Word:page numbers" \r "idxWordPgNos"

Linking is enabled by named ranges following a \r switch. Linked indexes in Word are created by third party applications that use the embedded XE fields to create an index that links locators to the referenced range names. XE fields become visible when hidden text is showing. Range names can be inspected using a bookmarks list.

Hovering the cursor over a linked locator displays the hyperlink. Following the hyperlink moves the cursor to the beginning of the range. Opening the hyperlink highlights the range.

Behaviour of a Linked Index in a PDF using Adobe Acrobat Pro

When a linked locator in the PDF is followed the beginning of the referenced range is displayed at the top of the screen.

Hyperlink details are not available via the locator, and the range name is not available when selected.

Survey

It was known that platforms and software behaved differently depending on configuration, but there was no clear picture of what the differences were. Indexers were asked to complete a simple survey to try and establish an overview, The results are shown below. No analysis is offered as some results are surprising, preventing a clear conclusion. Readers wanting to test their own systems are invited to test the links below, this will give a much clearer and faster result than reading the survey results which can only be a guide.

The IndexExploit user guide was used as the test document. Page numbers and bookmarks are known and in contains a linked index. The first hyperlink uses a sub address to link to page 14, the second uses a sub address to link to a bookmark on the same page. These links were chosen because they were used to answer a query about the formatting of locators when a Word embedded index is generated.

An initial survey conducted between 31st March and 10th April 2023 using the two links below enabled an analysis of behaviour of hardware and software combinations via the Sky and IDG indexing groups. The survey was continued, with an option of using the same links in a Word file and a PDF, more representative of how indexes will be served to users.

https://0201.nccdn.net/1_2/000/000/140/749/indexexploit-user-guide.pdf#page=14

https://0201.nccdn.net/1_2/000/000/140/749/indexexploit-user-guide.pdf#idxWordPgNos

Initial results from the survey showed these two links to be extremely successful tests. The results are shown in Table 1, Table 2 and Table 3.

Conclusions

Time will be required to assimilate and understand the results in the context of existing and future indexing challenges, the implications for digital publishing, and the desire by users to accurately +reference material for their own needs.

Some of the results are surprising. In some cases, the user is not taken to the correct page even though the configuration appears to be the same as others that are taken to the correct page.

There appears to be better support for page number sub addresses than for range names.

Microsoft Edge, Acrobat Pro and Word 365 (Desktop) performed well with range name and page sub addresses. The Apple MacBook recognised page and range name sub addresses with Safari. iPads didn't take users to the page. A Samsung Galaxy S8 smartphone took the user to the page sub address with Firefox, although Firefox doesn't recognise range name sub addresses.

Requirements

The knowledge gained from the survey, the application of IndexExploit, Mary Coe's presentation on the revised Kindle format to the 2023 ASI conference (Coe 2023) and other sources is used to define a set of requirements for linked locators intended to remove the inconsistencies uncovered by this survey, and to enable linking from multiple user applications.

Linked locators are used within electronic publications to link internally within a single document. Standalone indexes link to single or multiple online documents. The requirements below apply to all cases. The intention is to present a similar look and feel independent of the type of index, hardware, and software. The worldwide community of indexers and index users will benefit. Students and professionals will get improved access to online material. Organisations with large holdings of curated indexed documents will be able to improve access for users.

Definitions

Shall A mandatory requirement. As this document cannot mandate any requirements, shall is not used here.

Should A recommended requirement.

Will Something that can be relied on.

May/may not An optional state or requirement.

Must A statutory requirement such as government regulations. Considered beyond the scope of this document.

Hyperlinks Containing Page Number Sub Addresses

When a hyperlink containing a page number sub address is opened, the user **should** be taken to the correct page.

Page number links from a standalone index or document **may not** include the page number seen by the user but **will** be adjusted during creation for document structure.

Hyperlinks Containing Range Name Sub Addresses

When a hyperlink containing a range name sub address is opened, the user **should** be taken to the correct range.

Ranges and their names **will** be defined during document creation, which includes indexing.

A selected range **should** be made easily visible.

Where a selected range is small, it **should** be artificially extended to make it visible.

Users **should** be able to read range names and to determine their start and end locations.

Navigation

Contents lists, when present, **should** be easily accessible.

Indexes, where present, **should** be easily accessible.

Hyperlinks and Indexes

Barry Campbell

Draft C May 2023

Each index group A, B, C. etc. **should** be easily reachable.

Hyperlinks and Indexes

Draft C May 2023

Barry Campbell

Phase 1 results. Users test the links in an email message.

For this part of the survey the two hyperlinks were sent in an email. The results of attempting to open the hyperlinks are shown below. I've also included tests I ran from Word, Excel, PowerPoint, and Acrobat Pro.

The 'right' result is page 14.

Table 1 Phase 1 Survey Results

Date	Platform	Operating System	Software	Hyperlink result	
				Using a page sub address #page=14	Using a bookmark sub address #idxWordPgNos
31/3/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Office 365 Outlook running on Microsoft Edge	Page 14	Page 14
31/3/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Office 365 Word	Page 14	Page 14
31/3/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Office 365 Excel	Page 14	Page 14
31/3/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Office 365 PowerPoint	Page 14	Page 14
31/3/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Acrobat Pro	Page 14 Security warning on first attempt	Page 14
09/04/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Mozilla Firefox version 111.0.1	Page 14	Page 1. If tried immediately after #page=14, then the page is unchanged.
31/3/2023	iPad Air 5 th Gen	iOS 16.3.1	Office 365 Outlook running on Safari	Page 1	Page 1
9/4/2023	iPad Air 5 th Gen	iOS 16.3.1	Firefox 111.2	Page 1	Page 1
31/3/2023	Samsung Galaxy S8 phone (SM-G950F)	Android Version 9	Outlook Samsung Internet	Asks to download file	Asks to download file

Hyperlinks and Indexes

Draft C May 2023

Barry Campbell

Date	Platform	Operating System	Software	Hyperlink result	
9/4/2024	Samsung Galaxy S8 phone (SM-G950F)	Android Version 9	Link copied to Firefox 111.1.1 for Android	Page 14	Page 1
31/3/2023	Dell Inspiron desktop with Intel(R) Core(TM) i7-8700 processor	Windows 11 Pro	Microsoft Edge version 111.0.1661.54	Page 14	Page 14
31/3/2023	Dell OptiPlex 3020 (upgraded) with Intel(R) Core(TM) i5-4590 CPU	Windows 10 Pro	Firefox v. 111.0.1	Page 14	Page 1
31/3/2023	Dell OptiPlex 3020 (upgraded) with Intel(R) Core(TM) i5-4590 CPU	Windows 10 Pro	Microsoft Edge v.112.0.1722.34	Page 14	Page 14
31/3/2023	Dell OptiPlex 7070	Windows 10 Pro	Firefox 111.0.1 (64 bit)	Page 1	Page 1
4/4/2023	Acer Aspire desktop	Windows 11 Home	Firefox 111.0.1 (64 bit)	Page 14	Page 1
4/4/2023	Apple MacBook Air (13 in. 2017) WITH 2.2 GHz Dual-Core Intel Core i7	Monterey 12.6.3	Apple Mail/Safari	Page 14	Page 14
14/4/2023	ASUS ZenBook, 64-bit operating system, x64-based processor	Win 10 Pro 22H2, Windows Feature Experience Pack 120.2212.4190.0	Firefox 111.0.1	Page 14	Page 1
14/4/2023	PC	Windows 10 Pro	Chrome	Page 14	Page 1
14/4/2023	PC	Windows 10 Pro	Firefox	Page 14	Page 1
14/4/2023	PC	Windows 10 Pro	Edge	Page 14	Page 14
18/4/2023	PC	Windows 10	Windows Mail Mozilla Firefox 112.01	Page 1	Page 1

Phase 2a. Users test the links from a Word document

For the second part of the survey, users were sent a Word file and a PDF containing the hyperlinks. The results of attempting to open the hyperlinks are shown below.

Hyperlinks and Indexes

Draft C May 2023

Barry Campbell

Table 2 Phase 2a Survey Results for the Word file

Date	Platform	Operating System	Software	Hyperlink result	
				Using a page sub address #page=14	Using a bookmark sub address #idxWordPgNos
12/4/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home (KB5023706)	Office 365 Word	Page 14	Page 14
11/4/2023	Dell OptiPlex 7070	Windows 10 Pro	Word 2016	Page 1	Page 1
13/4/2023	Dell OptiPlex 3020 (upgraded) with Intel(R) Core(TM) i5-4590 CPU	Windows 10 Pro	Word 2010	Page 14	Page 1
13/4/2023	ASUS ZenBook laptop	Windows 10	Open Office Write	No follow	No follow
13/4/2023	ASUS ZenBook laptop	Windows 10	WPS Office	Page 1	No follow
14/4/2023	MacBook Pro, early 2015 3.1 GHz Dual-Core Intel Core i7	macOS Monterey (12.6.3)	LibreOffice Writer, Firefox	Page 14	Page 1
14/4/2023	Apple MacBook Air (13 in. 2017) WITH 2.2 GHz Dual-Core Intel Core i7	Monterey 12.6.3	Word (v. 16.72)	Page 14	Page 1
14/4/23	Dell XPS laptop (XPS 15 9520), Intel I-7 processor	Windows 11 Home	email via Microsoft Edge browser and internet webmail interface. Downloaded Word file to laptop, opened it in Word (Office 2021).	Page 14	Page 14
14/4/23	Dell XPS laptop (XPS 15 9520), Intel I-7 processor	Windows 11 Home	email via Microsoft Edge browser and internet webmail interface. Downloaded Word file to laptop, opened it in Word (Office 2021).	Page 14	Page 14

Phase 2b Users test links from a PDF

Hyperlinks and Indexes

Draft C May 2023

Barry Campbell

Table 3 Phase 2b Survey Results for the PDF File

Date	Platform	Operating System	Software	Hyperlink result	
				Using a page sub address #page=14	Using a bookmark sub address #idxWordPgNos
12/4/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Acrobat Pro	Page 14	Page 14
12/4/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Firefox	Page 14	Page 1
12/4/2023	HP Envy 11 th Gen Intel Core i7	Windows 11 Home	Edge	Page 14	Page 14
11/4/2023	Dell OptiPlex 7070	Windows 10 Pro	Adobe Acrobat X Pro	Page 1	Page 1
13/4/2023	Dell OptiPlex 3020 (upgraded) with Intel(R) Core(TM) i5-4590 CPU	Windows 10 Pro	Firefox v. 112.0	Page 14	Page 1
13/4/2023	Dell OptiPlex 3020 (upgraded) with Intel(R) Core(TM) i5-4590 CPU	Windows 10 Pro	Edge	Page 14	Page 14
13/4/2023	ASUS ZenBook laptop	Windows 10	Foxit PDF reader	Page 14 after security pop-up	Page 1 after security pop-up
13/4/2023	ASUS ZenBook laptop	Windows 10	Adobe Acrobat Reader	Page 14 after security pop-up	Page 1 after security pop-up
14/4/2023	MacBook Pro, early 2015 3.1 GHz Dual-Core Intel Core i7	macOS Monterey (12.6.3)	PDF Expert	Page 14	Page 1
14/4/2023	Apple MacBook Air (13 in. 2017) WITH 2.2 GHz Dual-Core Intel Core i7	Monterey 12.6.3	Apple Preview (v. 11.0 (1033.4))	Page 14	Page 1
14/4/23	Dell XPS laptop (XPS 15 9520), Intel I-7 processor	Windows 11 Home	Microsoft Edge	Page 14	Page 14

Bibliography

- CERN. 2023. *A Short History of the Web*. 4th March. <https://home.cern/science/computing/birth-web/short-history-web>.
- Coe, Mary. 2023. "E-book Indexes:More than Meets the Eye." *ASI Conference 2023*.
- Hoffman, Jay. 2023. *The History of the Web*. Accessed April 24, 2023. <https://thehistoryoftheweb.com/timeline/>.
- . 2021. *The Right to Link*. 18 May. Accessed April 24, 2023. <https://thehistoryoftheweb.com/the-right-to-link/>.
- IMDb. 2023. *IMDb*. Accessed April 24, 2023. <https://www.imdb.com/>.
- IndexBase. 2023. *IndexExploit Downloads*. Accessed May 01, 2023. <http://www.indexbase.co.uk/indexexploit/indexexploitdownloads>.
- Jehoram, Tobias Cohen. 2016. *European Court of Justice: hyperlinks to unauthorized content may infringe copyright*. December. Accessed April 24, 2023. https://www.wipo.int/wipo_magazine/en/2016/06/article_0007.html.
- Ream, Dave. 2016. *A Comparative Look at Embedded Index Entry*. October. Accessed April 25, 2023. <https://digital-publications-indexing.org/wp-content/uploads/2017/10/Embedded-Index-Entry-Coding-Ream.pdf>.
- W3C. 2023. *Fragments*. Accessed April 24, 2023. https://www.w3.org/Addressing/URL/4_2_Fragments.html.
- Wikipedia. 2023. *Hypertext Fiction*. Accessed April 24, 2023. https://en.wikipedia.org/wiki/Hypertext_fiction.

Barry Campbell is an indexer based in the UK. He's the developer of IndexExploit for Word embedded indexing, linking, and hyperlink manipulation; and of IndexConvert for importing indexes into indexing software. In a previous career he gained experience of a wide range of requirement management tools and was responsible for requirements management of large aerospace and defence projects.

email barry.campbell@indexbase.co.uk

website www.indexbase.co.uk