



BCNET
CONNECT
HIGHER ED & RESEARCH TECH SUMMIT

Eliminating technical debt in Digital Humanities projects

Janelle Jenstad and Martin Holmes
University of Victoria Humanities Computing and Media Centre

What is HCMC?

- Humanities Computing and Media Centre
- Set up in the mid-1990s
- Provides programming support for research and teaching in the Humanities
- Includes a lab where researchers and teams can work



Eliminating technical debt in DH projects

What is HCMC?

- Has published hundreds of digital resources over the years
- Many of these are large-scale digital editions, such as:
 - The Map of Early Modern London (13,257 pages, mapoflondon.uvic.ca)
 - The Colonial Despatches of BC and Vancouver Island (10,827 pages, bcgenesis.uvic.ca)
 - The Landscapes of Injustice Archive (154,846 pages, loi.uvic.ca/archive)

The 1990s

- We made websites mostly to support language teaching.
- We build lots of interactive exercises using HTML, CSS and JavaScript.
- Most of these sites have never been updated – see for example:
- web.uvic.ca/hrd/indonesian/



B E G I N N I N G I N D O N E S I A N

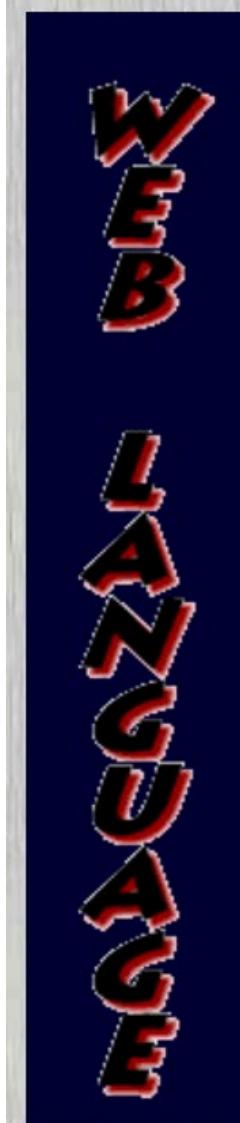
Explanation in English	About Indonesian at UVic
Introductory Unit (Try it out!)	Glossary
Unit 1	Unit 8
Unit 2	Unit 9
Unit 3	Unit 10
Unit 4	Unit 11
Unit 5	Unit 12
Unit 6	Unit 13
Unit 7	Unit 14
Other Indonesian sites	Send us your comments



The 1990s

- These sites all still work – even though they look SOOO 1990s:
- web.uvic.ca/lancenrd/martin/weblang/

No maintenance burden from these sites!



Welcome!

Before you begin this presentation, please note that you will need to be using version 3 or above of Netscape Navigator or Microsoft Internet Explorer, and JavaScript should be enabled in your browser. Click the button below to find out whether your browser is set up correctly.

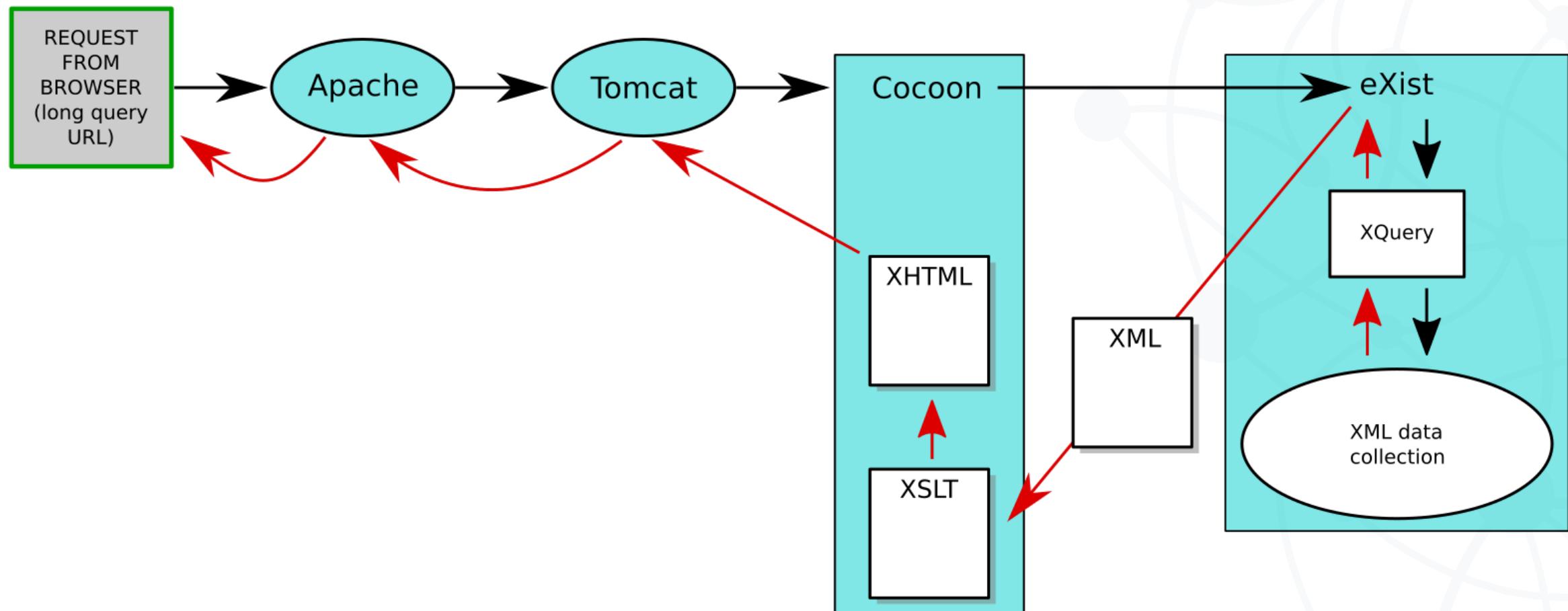
[Click here](#)

If nothing happens when you click the button, you will need to enable JavaScript in the Options of your browser.

[Next page](#)

The early 2000s

- Digital Humanities became a thing.
- New technologies came along.
- We got ambitious and enthusiastic, and started building larger, more complex sites with:
 - ASP
 - MySQL, PostgreSQL, PHP
 - Apache Cocoon
 - The eXist XML database
 - ...and lots of other cool new tools.



Apparent advantages of this sort of stack

- Changes to output rendering propagate instantly to the whole collection.
- Individual documents can be published as soon as they're ready ("rolling publication").

BUT: Horrible technical debt

A dense word cloud composed of various programming languages and frameworks. The most prominent words are Node, PHP, Ruby, .NET, Python, MySQL, Tomcat, JSP, Cocoon, React, PostgreSQL, MongoDB, Django, Angular, jQuery, Spring, and WordPress. Other visible words include exist, ASPX, Vue, and React.

BUT: Horrible technical debt

Update Java: Tomcat breaks.

Update Tomcat: Cocoon breaks.

Update Cocoon: eXist breaks.

Update eXist: XQuery has to be rewritten.

...

OS update requires MySQL update.

MySQL update requires PHP update.

PHP update requires code rewrite.

...

Miss an update and your site is vulnerable to attack.

...



Eliminating technical debt in DH projects

Now every project is a Tamagotchi.

Life is nothing but maintenance.

There's no time for new projects.

Eliminating technical debt in DH projects



But wait – it's worse. Remember this?

- Individual documents can be published as soon as they're ready (“rolling publication”).

It seemed like a good idea at the time...

The trouble is, there's no such thing as a document.

The screenshot shows a detailed view of the "RAM ALLEY" page on the "Map of Early Modern London" website. The header features a decorative map background, navigation links for "Map", "Encyclopedia", "Library", "Stow", "About", and "News", and a search bar with a "Go!" button. The main content area has a breadcrumb trail: "Home > Encyclopedia > Placeography > Streets".

RAM ALLEY

- * [Agas Map](#)
- * [Google Map](#)
- * [List documents mentioning Ram Alley](#)
- * [List variant names and spellings](#)

LOCATION

Ram Alley, a mere seven feet wide, ran southwards from Fleet Street, opposite Fetter Lane. Its end point was a footway between two legal institutions: the Inner Temple and Serjeants Inn. Edward H. Sugden also mentions that the street was well known as the rear exit from another inn, the **Mitre**, which fronted onto Fleet Street.

ETYMOLOGY

The alley was named after an inn, marked by the sign of the **Star and Ram**, which had originally belonged to the Knights Hospitallers but was confiscated by Henry VIII. It was taken in fee from the monarch for £54 by Robert Harrys, or Harris, and became the site of his brewery, which had a frontage on Fleet Street (Bell 247). The alley is now known as **Hare Place**, named after Hare House (Paige letter #154).

MAP VIEWS

An unlabelled alley in the correct location as **Ram Alley** appears on the [Agas map](#). The alley is marked on both the [Ogilby and Morgan map of 1676](#) and the [Rocque map of 1746](#).

The trouble is, there's no such thing as a document.

The screenshot shows a web page from the "Map of Early Modern London" website. The header features a decorative background map of London and includes links for "Map", "Encyclopedia", "Library", "Stow", "About", and "News". A search bar with a "Go!" button is also present. The main content area has a breadcrumb navigation path: "Home > Encyclopedia > Placeography > Streets". The title of the page is "RAM ALLEY". Below the title is a red triangular graphic. The page contains several sections with links:

- * Agas Map
- * Google Map
- * List documents mentioning Ram Alley
- * List variant names and spellings

LOCATION

Ram Alley, a mere seven feet wide, ran southwards from Fleet Street, opposite Fetter Lane. Its end point was a footway between two legal institutions: the Inner Temple and Serjeants Inn. Edward H. Sugden also mentions that the street was well known as the rear exit from another inn, the Mitre, which fronted onto Fleet Street.

ETYMOLOGY

The alley was named after an inn, marked by the sign of the Star and Ram, which had originally belonged to the Knights Hospitallers but was confiscated by Henry VIII. It was taken in fee from the monarch for £54 by Robert Harrys, or Harris, and became the site of his brewery, which had a frontage on Fleet Street (Bell 247). The alley is now known as Hare Place, named after Hare House (Paige letter #154).

MAP VIEWS

An unlabelled alley in the correct location as Ram Alley appears on the Agas map². The alley is marked on both the Ogilby and Morgan map of 1676 and the Rocque map of 1746.

The trouble is, there's no such thing as a document.

The screenshot shows a web page from the "Map of Early Modern London" website. The header features a decorative background map of London and includes links for "Map", "Encyclopedia", "Library", "Stow", "About", and "News". A search bar with a "Go!" button is also present. The main content area has a breadcrumb navigation path: "Home > Encyclopedia > Placeography > Streets". The page title is "RAM ALLEY" with a red triangle icon containing "AJAX". Below the title are several links: "Agas Map", "Google Map", "List documents mentioning Ram Alley", and "List variant names and spellings". The "LOCATION" section contains a paragraph about Ram Alley's history, mentioning its width, location between legal institutions, and its connection to Fleet Street. The "ETYMOLOGY" section discusses the name's origin, linking it to an inn called "Star and Ram". The "MAP VIEWS" section notes the alley's presence on historical maps by Agas, Ogilby, and Rocque. The left sidebar contains navigation links for "Page contents", "Credits", and author information for Jacqueline Watson, Tye Landels, Joey Takeda, and Joey Takeda again.

The trouble is, there's no such thing as a document.

Map Encyclopedia Library Stow About News

Go!

MAP of EARLY MODERN LONDON

Page contents Credits

Author:
Jacqueline Watson

Encoder:
Tye Landels

Copy Editor:
Tye Landels

Author of Abstract:
Joey Takeda

Geographic Information Specialist:
Joey Takeda

Home > Encyclopedia > Placeography > Streets

RAM ALLEY

- * Agas Map
- * Google Map
- * List documents mentioning Ram Alley
- * List variant names and spellings

LOCATION

Ram Alley, a mere seven feet wide, ran southwards from Fleet Street, opposite Fetter Lane. Its end point was a footway between two legal institutions: the Inner Temple and Serjeants Inn. Edward H. Sugden also mentions that the street was well known as the rear exit from another inn, the Mitre, which fronted onto Fleet Street.

ETYMOLOGY

The alley was named after an inn, marked by the sign of the Star and Ram, which had originally belonged to the Knights Hospitallers but was confiscated by Henry VIII. It was taken in fee from the monarch for £54 by Robert Harrys, or Harris, and became the site of his brewery, which had a frontage on Fleet Street (Bell 247). The alley is now known as Hare Place, named after Hare House (Paige letter #154).

MAP VIEWS

An unlabelled alley in the correct location as Ram Alley appears on the Agas map². The alley is marked on both the Ogilby and Morgan map of 1676 and the Rocque map of 1746.

The trouble is, there's no such thing as a document.

Map Encyclopedia Library Stow About News

Go!

MAP of
EARLY MODERN
LONDON

Page contents

Credits

Author:
Jacqueline Watson

Encoder:
Tye Landels

Copy Editor:
Tye Landels

Author of Abstract:
Joey Takeda

Geographic Information
Joey Takeda

Home > Encyclopedia > Placeography > Streets

RAM ALLEY

- * Agas Map
- * Google Map
- * List documents mentioning Ram Alley
- * List variant names and spellings

LOCATION

Ram Alley, a mere seven feet wide, ran southwards from Fleet Street, opposite Fetter Lane. Its end point was a footway between two legal institutions: the Inner Temple and Serjeants Inn. Edward H. Sugden also mentions that the street was well known as the rear exit from another inn, the Mitre, which fronted onto Fleet Street.

ETYMOLOGY

The alley was named after an inn, marked by the sign of the Star and Ram, which had originally belonged to the Knights Hospitallers but was confiscated by Henry VIII. It was taken in fee from the monarch for £54 by Robert Harrys, or Harris, and became the site of his brewery, which had a frontage on Fleet Street (Bell 247). The alley is now known as Hare Place, named after Hare House (Paige letter #154).

MAP VIEWS

An unlabelled alley in the correct location as Ram Alley appears on the Agas map². The alley is marked on both the Ogilby and Morgan map of 1676 and the Rocque map of 1746.

The trouble is, there's no such thing as a document.

The screenshot shows a web page from the "Map of Early Modern London" website. The page title is "RAM ALLEY". On the left sidebar, under "Page contents", there are several red triangles pointing towards the main content area, each containing the word "AJAX". The main content area includes sections for "LOCATION" and "ETYMOLOGY", both of which are also heavily annotated with red triangles containing "AJAX". The "LOCATION" section contains a paragraph of text. The "ETYMOLOGY" section contains a longer paragraph of text. At the bottom, there is a section titled "MAP VIEWS" with a note about unlabelled alleys appearing on historical maps.

Map Encyclopedia Library Stow About News

Go!

MAP of
EARLY MODERN
LONDON

Page contents

Credits

Author:
Jacqueline Watson

Encoder:
Tye Landels

Copy Editor:
Tye Landels

Author of Abstract:
Joey Takeda

Geographic Information:
Joey Takeda

RAM ALLEY

- * Agas Map
- * Google Map
- * List documents mentioning Ram Alley
- * List variant names and spellings

LOCATION

Ram Alley, a mere seven feet wide, ran southward between two legal institutions: the Inner Temple and Seven Stars Inn. Edward H. Sugden has suggested that the street was well known as the rear exit from another inn, the Mitre, which fronted onto Fleet Street.

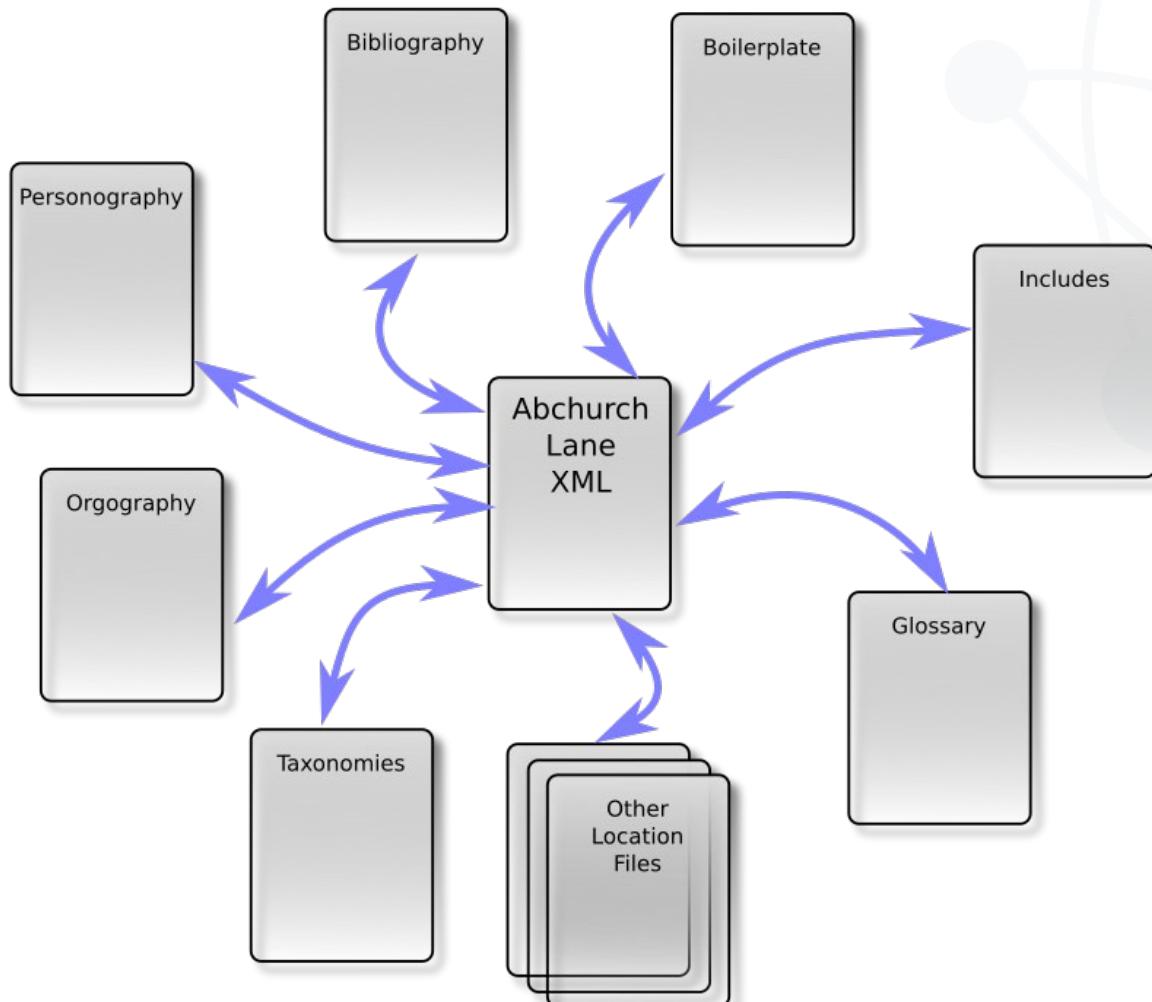
ETYMOLOGY

The alley was originally called Hare Place, marked by the sign of a Hare and Ram, which had belonged to the Knights Hospitaller. It was confiscated by Henry VIII and taken in fee from the Knights for £54 by Robert Harris, or Harris, and became the site of his building. It had a frontage on Fleet Street (Bell 247). The alley is now known as Hare Place, named after Hare House (Paige letter #154).

MAP VIEWS

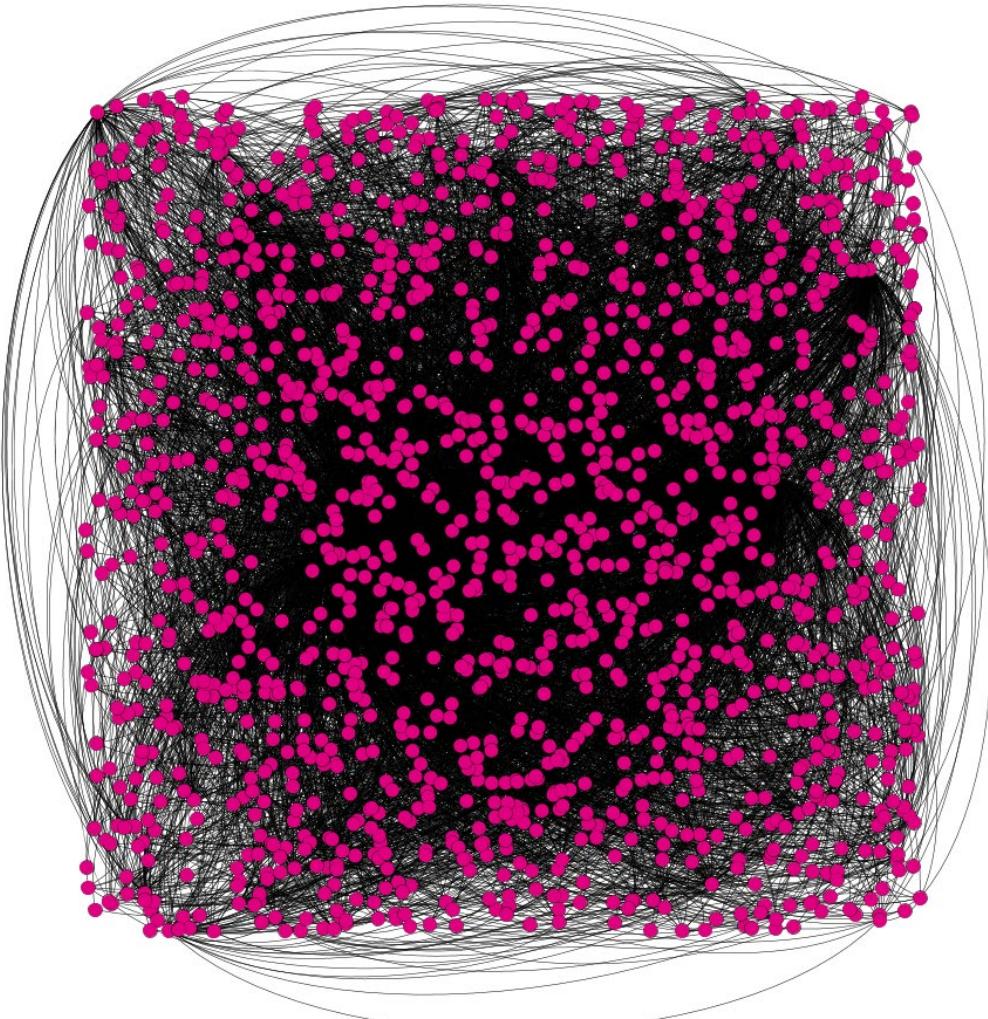
An unlabelled alley in the correct location as Ram Alley appears on the Agas map². The alley is marked on both the Ogilby and Morgan map of 1676 and the Rocque map of 1746.

A digital edition is a network, not a collection of discrete docs.



Eliminating technical debt in DH projects

Network map of MoEML



Eliminating technical debt in DH projects

The scholarly corollary

- Changes to one “document” cause changes to other “documents”.
- There is no reliable way to cite a “page”.
- There’s no way to get back to the version of a “page” you saw last week.
- So nobody will cite the project.
- So the scholars involved will not win prizes/get tenure/get grant money/be happy and successful.

So the Tamagotchi model is:

- bad for developers because they spend their lives on maintenance; and
- bad for scholars because their work appears untrustworthy and uncitable.
- So let's stop thinking like software developers, and start thinking like proper humanists.

What books got right

German *Edition* (1537)

Dutch *editie* (1546)

English *edition* (1556)

Italian *edizione* (1575)

“edition”

Edmund Coote, *The English Schoolmaster* (1596): “putting forth”

John Florio, *A World of Words* (1598): “a setting foorth, a publishing, a creating or an edition”

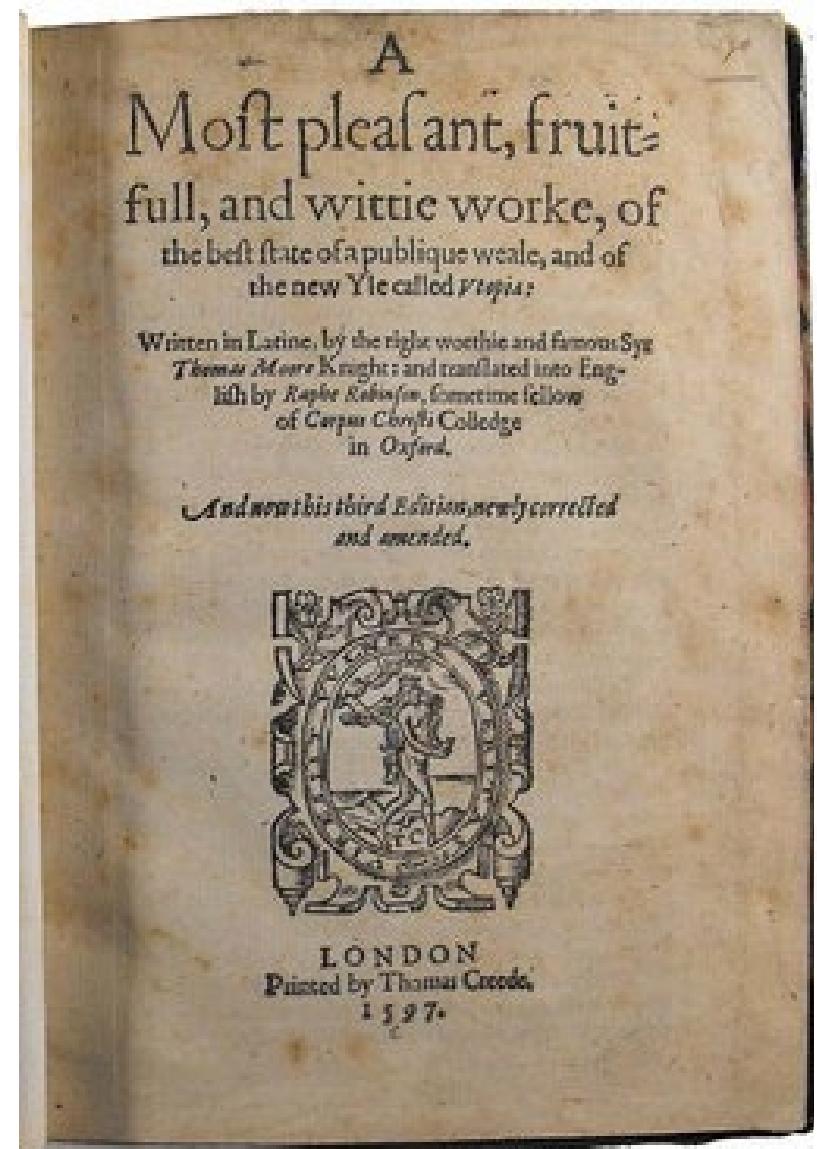
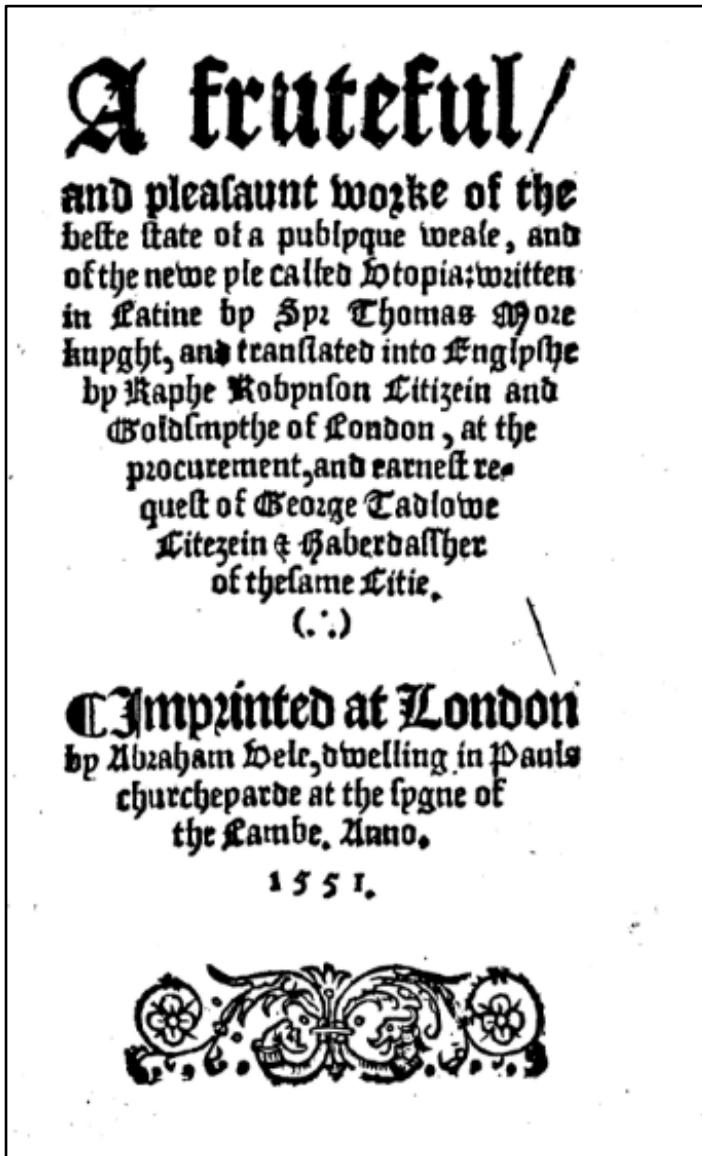
Robert Cawdrey, *A Table Alphabetical* (1604): a “putting fourth, setting abroade”

Utopia

By Thomas More



1516



New Edition as New Text

I haue now in this seconde edition taken about it such paines, yt verye fewe great faultes & notable errorrs are in it to be founde.

– Ralphe Robynson, second edition of his translation of
Thomas More's *Utopia*, 1556 (A3v)

New edition as text

A printed version of a given work that may be distinguished from other versions
... by its membership of a complete batch of copies printed from the same
setting of type, usually at the same time and place. These batches come to be
numbered as first, second, third, etc. editions, each time a new version is set
again from fresh type.

- Baldick, *Oxford Dictionary of Literary Terms*

Editions ≈ Versions



5.48 Version: What It Is

If the source carries a notation indicating that it is a version of a work released in more than one form, identify the version in your entry. Books are commonly issued in versions called *editions*. A revised version of a book may be labeled *revised edition* or be numbered (*second edition*, etc.). Versions of books are sometimes given other descriptions as well.

Rolling releases ≈ wikis

... you should never end your research with *Wikipedia* or another wiki. Instead, use such sites at the very beginning stages of your research to gather basic information and sources and then move on to more credible sources.

– Ellen C. Carillo, *MLA Guide to Digital Literacy*, 104

Wikis don't trust wikis

Wikipedia is not a reliable source for citations elsewhere on Wikipedia. As a user-generated source, it can be edited by anyone at any time, and any information it contains at a particular time could be vandalism, a work in progress, or simply incorrect.

https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_not_a_reliable_source

We claim ...

- Style guides won't help us cite a moving target.
- The onus is on producers of resources, not citers of resources.
- The digital resources themselves need to move to **stable** versioning/editioning.



The Endings Project

Building Sustainable Digital Humanities Projects



Endings Team and Questions

- Humanities faculty
 - Librarians
 - Programmers
-
- How do and how should DH projects conclude?
 - How should we preserve projects to retain their dynamic features?
 - Where should projects be archived?

Principles

Endings Principles for Digital Longevity
Version 2.2.1, 2023-03-03

We divide digital projects into five primary components:

1. Data
2. Documentation
3. Processing
4. Products
5. Release management

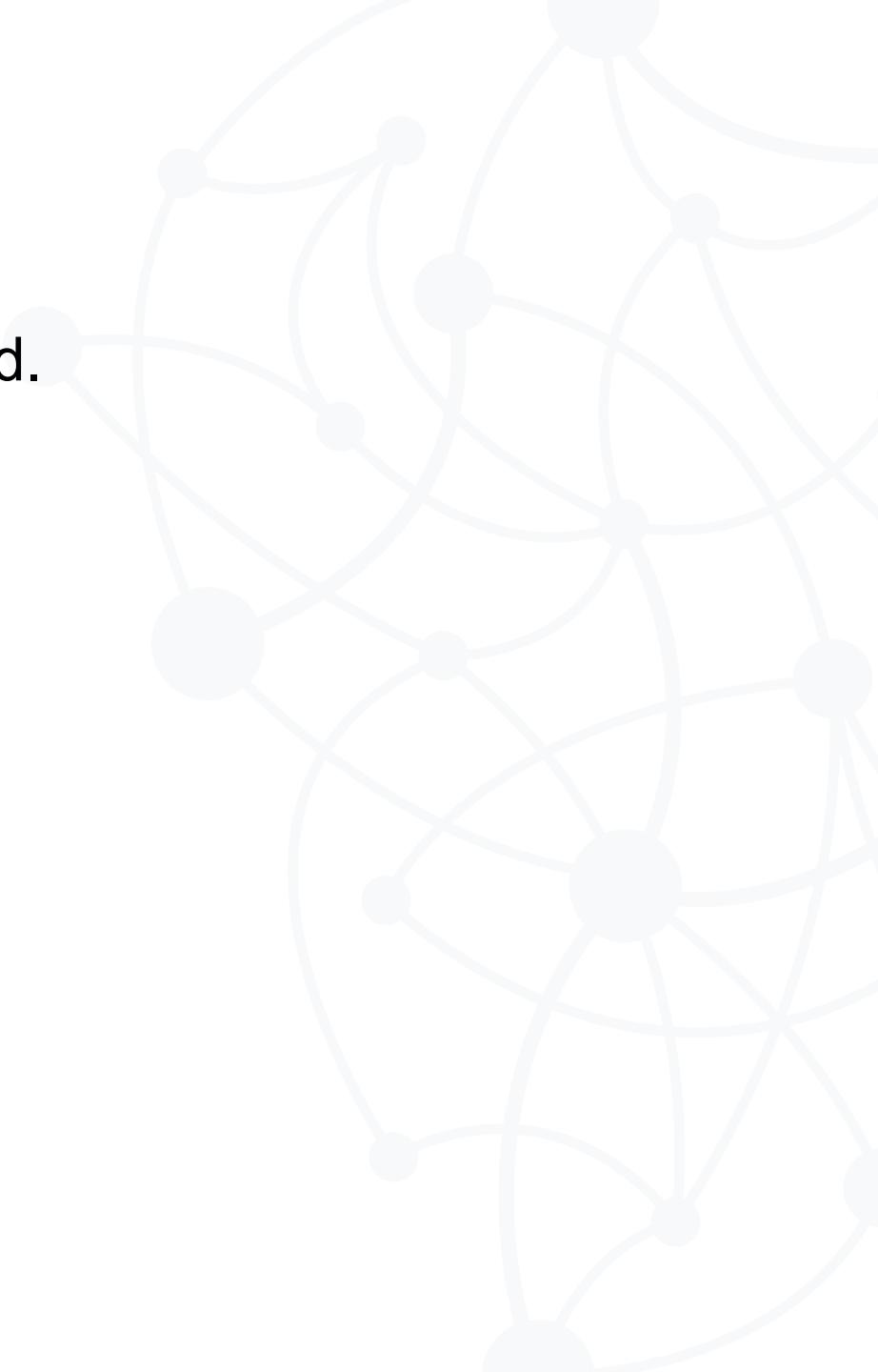
<https://endings.uvic.ca/principles.html>

Endings Principles 5: Release Management

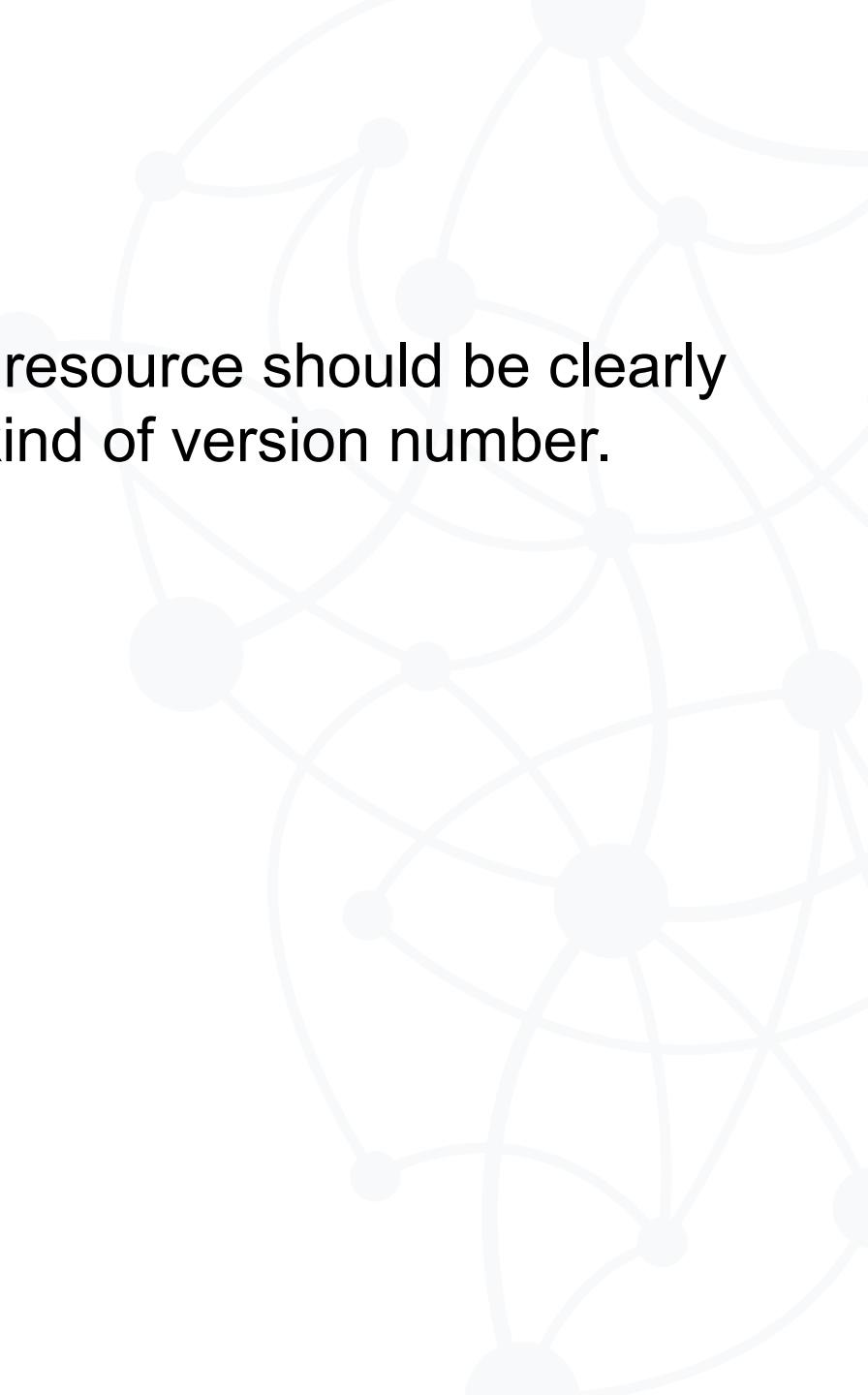
Release management handles the public release of products. Without good release management, a project can never end gracefully.

– <https://endings.uvic.ca/principles.html#release-management>

5.1 Releases should be **periodic** and carefully planned.



5.2 A release should only be made when the entire product set is **coherent, consistent, and complete** (passing all validation and diagnostic tests).



5.3 Like editions of print works, each release of a web resource should be clearly identified on every page by its build date and some kind of version number.

How MoEML does it

MoEML v.7.0, svn rev. 20565 2022-05-05 09:11:13 -0700 (Thu, 05 May 2022).

[Home](#) [Legal](#) [Teaching](#) [Cite MoEML](#) [Site map](#) [MoEML A-Z](#) [Contact us](#) [Donate](#)



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada

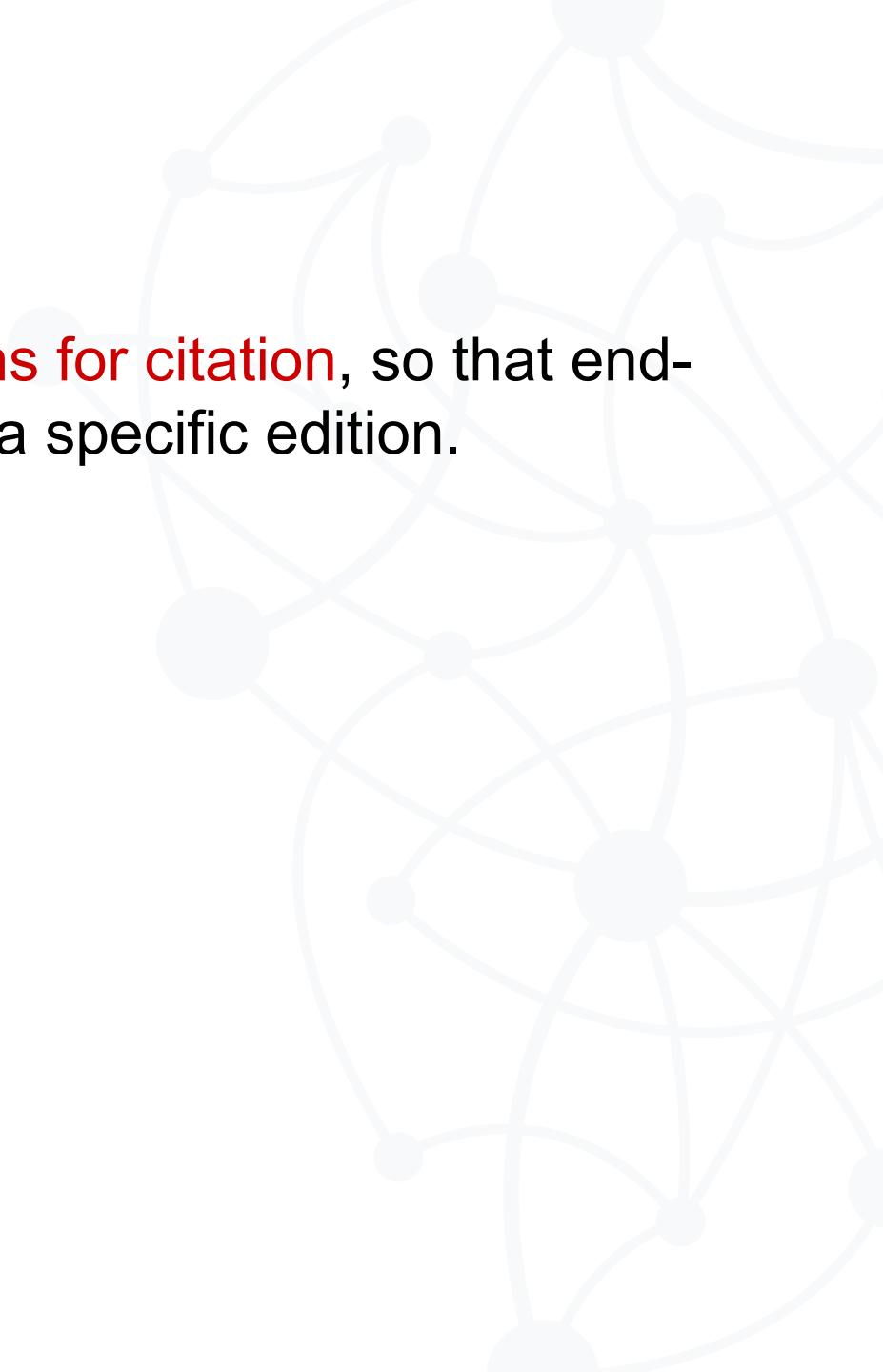
Canada



University
of Victoria

Humanities Computing
and Media Centre

MoEML v.7.0, svn rev. 20565 2022-05-05 09:11:13 -0700 (Thu, 05 May 2022).



5.4 Web resources should include detailed **instructions for citation**, so that end-users can unambiguously cite a specific page from a specific edition.

Cite this page

MLA citation

Middleton, Thomas. *The Triumphs of Truth. The Map of Early Modern London*, Edition 7.0, edited by **Janelle Jenstad**, U of Victoria, 05 May 2022, mapoflondon.uvic.ca/edition/7.0/TRIU1.htm.

Chicago citation

Middleton, Thomas. "The Triumphs of Truth." *The Map of Early Modern London*, Edition 7.0. Ed. **Janelle Jenstad**. Victoria: University of Victoria. Accessed May 05, 2022. mapoflondon.uvic.ca/edition/7.0/TRIU1.htm.

APA citation

Middleton, T. 2022. The Triumphs of Truth. In **J. Jenstad** (Ed), *The Map of Early Modern London* (Edition 7.0). Victoria: University of Victoria. Retrieved from <https://mapoflondon.uvic.ca/editions/7.0/TRIU1.htm>.

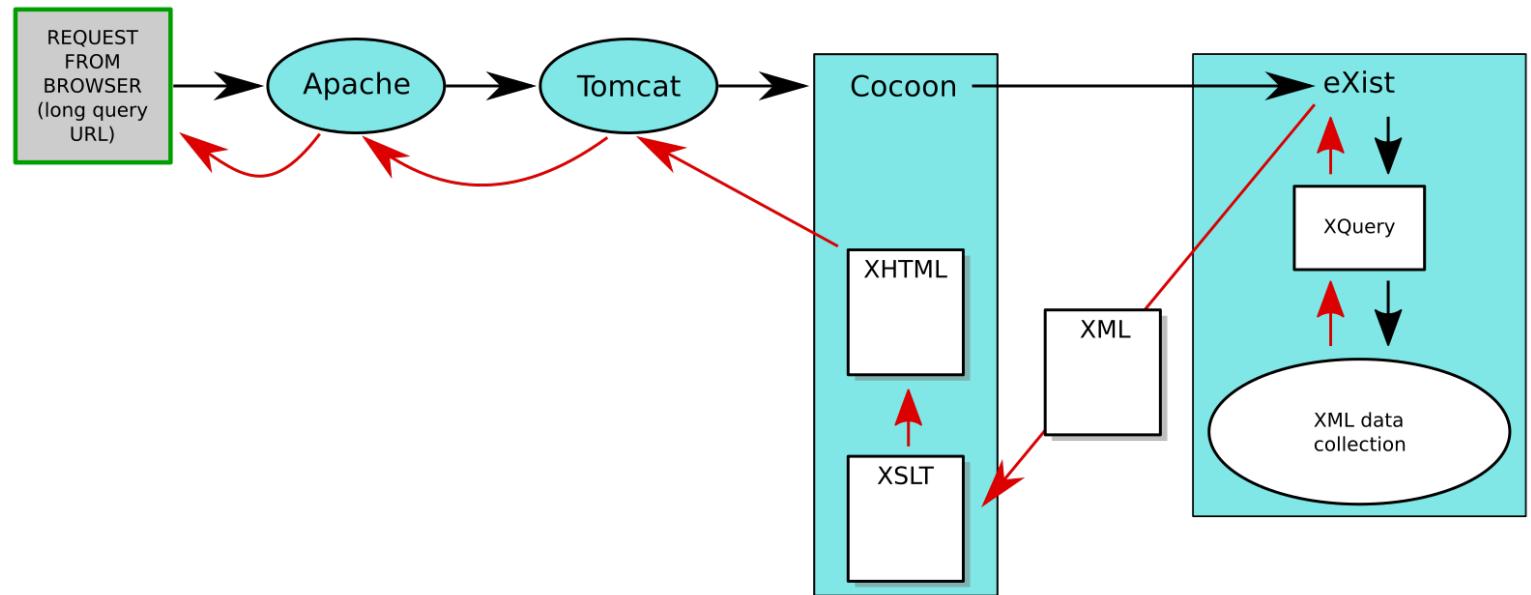
We want lots of static copies.

- Another desirable feature of the book is its durability.
- Books are durable partly because they're multiple and scattered.
- We need to do the same with a digital edition – make lots of copies.
- It's hard to make lots of copies of a complex database-driven infrastructure.
- Therefore we need our digital editions to be simple, static, portable, and easily copied.

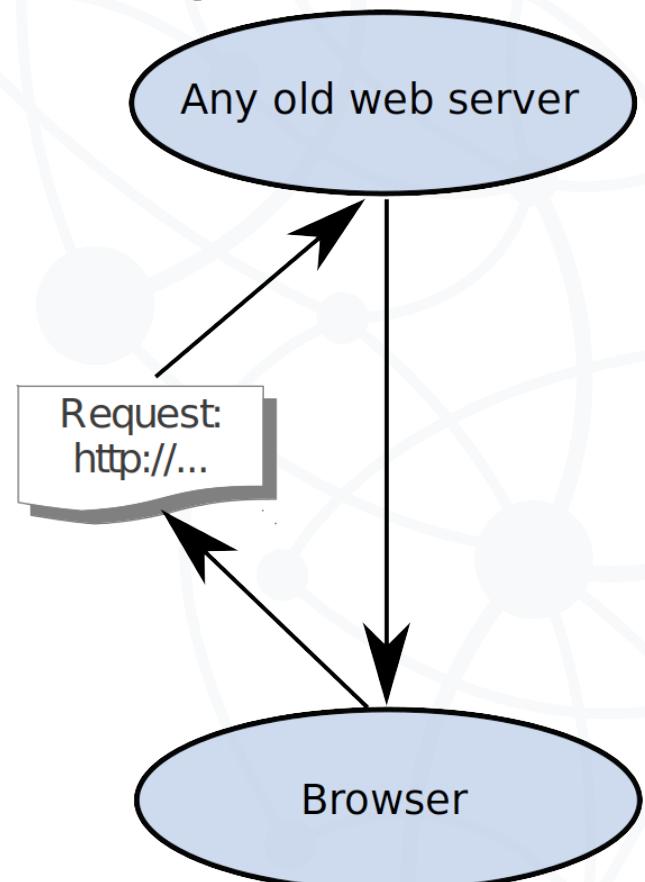
What is a static site?

- What we release: a static, numbered edition with no server-side dependencies
- HTML, JavaScript and CSS: every page is built at the same time in a single operation from a single source dataset.
- You can put it on thumb drives and distribute them to run on any web server.
- You can download it as a ZIP file and install it on many machines.
- The copy we host at the UVic URL is the canonical copy but it's just ONE of many static copies (like the public library copy).

Not this:



But this:



Endings Principles 4: Products (1)

- 4.1 No dependence on server-side software: build a static website with no databases, no PHP, no Python.
- 4.2 No boutique or fashionable technologies: use only standards with support across all platforms, whose long-term viability is assured. Our choices are **HTML5**, **JavaScript**, and **CSS**.
- 4.3 No dependence on external libraries or services: no JQuery, no AngularJS, no Bootstrap, no Google Search.

- So: *free of technical debt*.



In other words, a *pet rock*.



Endings Principles 4: Products (2)

- 4.4 **No query strings**: every entity in the site has a unique page with a simple URL that will function on any domain or IP address.
- http://graves.uvic.ca/xbrowse.xq;jsesessionid=926A438A2B364B4B92488B9AC00AB013?query_stored=true&collection=/db/graves/entries&type=diaryentry&n=1935-02-22&start=1

https://graves.uvic.ca/diary_1935-02-22.html

Endings Principles 4: Products (3)

- 4.5 **Inclusion of data**: every site should include a documented copy of the source data, so that users of the site can repurpose the work easily.
- 4.6 **Massive redundancy**: every page contains all the components it needs, so that it will function without the rest of the site if necessary, even though doing so means duplicating information across the site.
- 4.7 **Graceful failure**: every page should still function effectively even in the absence of JavaScript or CSS support.
 - So: *stable, citable, archivable, replicable, Wayback-Machinable, stealable, reusable*

How we build a static site (1)

Validate the **source XML** (using RNG and Schematron).

Create **generated XML** documents for indexes, TOCs, listings pages, etc.

Validate this new XML.

Create “**standalone**” versions of all the documents, by importing copies of all linked components (bios of people, place and org info, glossary entries, etc.)

Validate this standalone XML.

How we build a static site (2)

Build/copy CSS, JS, images, and other resources.

Create **XHTML5 output** from each of the standalone XML documents.

Validate the HTML using the **W3C validator**.

Run additional **diagnostics tests** and produce reports.

How do we do search on these sites?

- We wrote a staticSearch engine that does not require anything server-side except a web server:

<https://github.com/projectEndings/staticSearch>

It supports stemming, wildcards, phrasal searches, and metadata filters. Try it out:

- <https://bcgenesis.uvic.ca/search.html>



Endings **Compliance**

Deliberately coherent, consistent,
complete, versioned, valid, dated,
portable, static, no server scripting,
no databases



<https://endings.uvic.ca/compliance.html>

Summary

Digital editions must be citable with the precision and reliability of print editions.

Cited materials must persist and remain accessible without maintenance.

So:

For scholars, we must make coherent, consistent, complete, identifiable editions.

For developers, we must eliminate maintenance, technical debt, and vulnerability.

Thanks

The Endings Team: Claire Carlin, Ewa Czaykowska-Higgins, Janelle Jenstad; Stewart ArNeil, Martin Holmes, Greg Newton; John Durno, Matt Huculak, Lisa Goddard; Emily Comeau and Joey Takeda

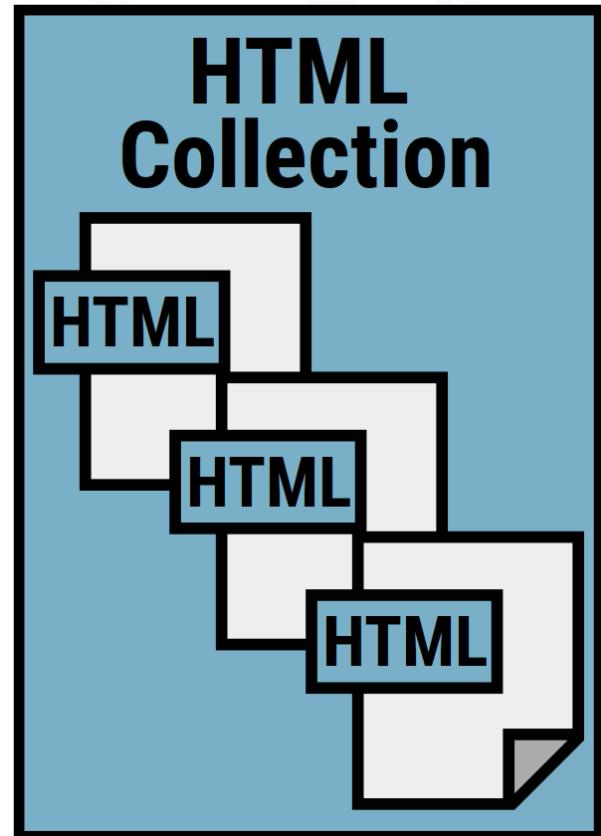
Faculty of Humanities
University of Victoria Libraries
Social Sciences and Research Council of Canada

SSHRC = CRSH

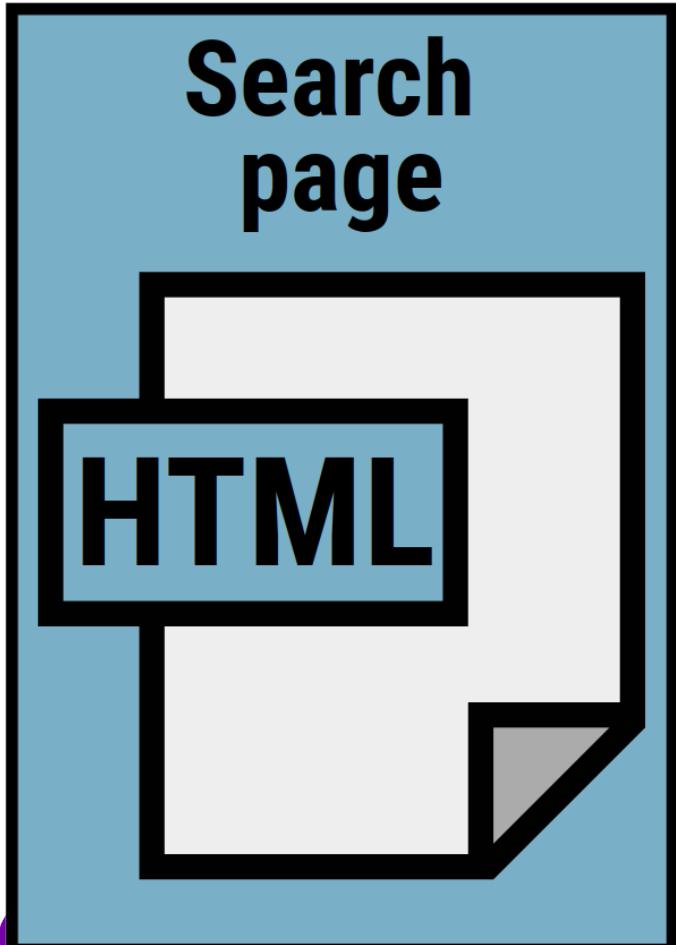
APPENDIX: How the staticSearch engine works

You have:

- A collection of HTML files
- All well-formed & valid XHTML



You add:



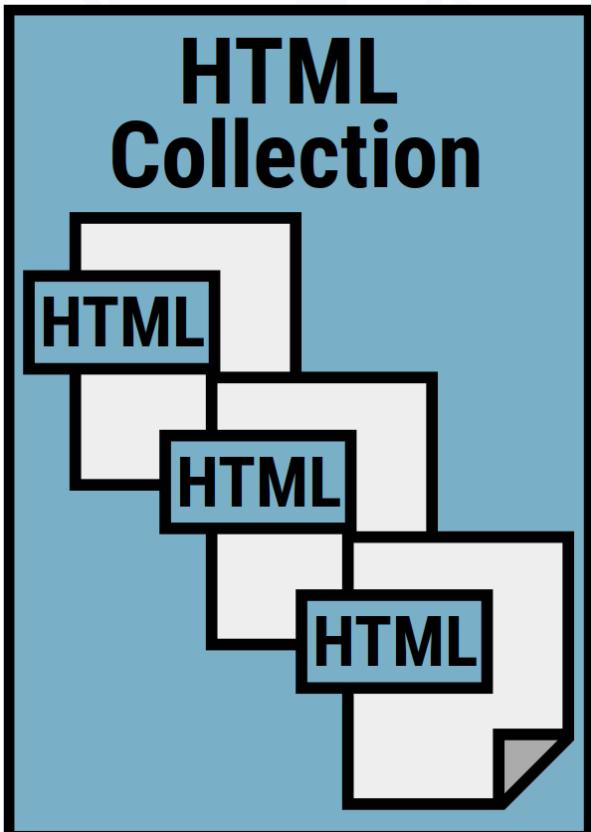
- An HTML search page containing a special div:

```
<div id="staticSearch"></div>
```

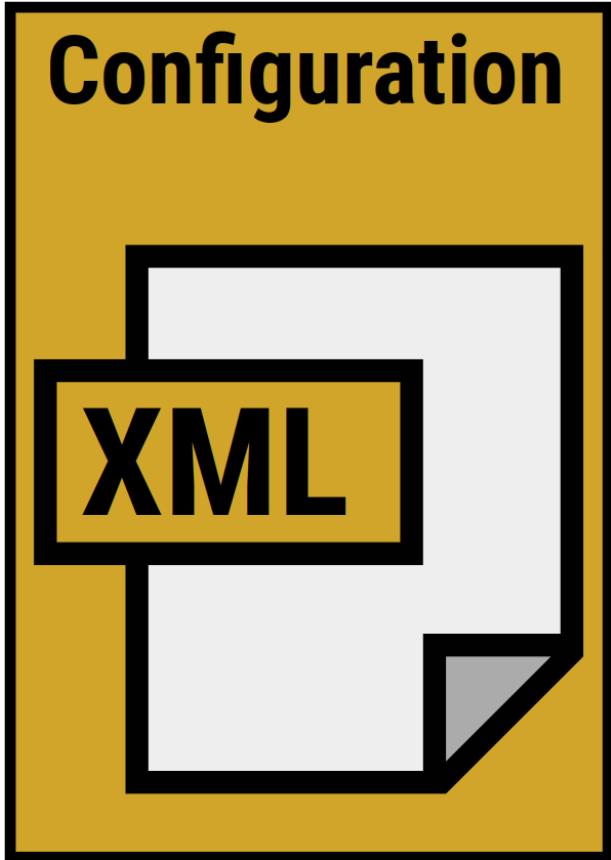
You add:

- Some metadata elements to support search filters:

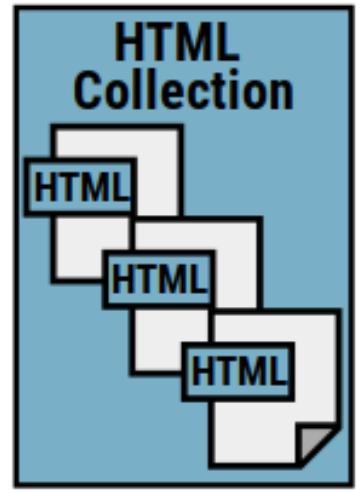
```
<meta name="Document type"  
      class="staticSearch.desc"  
      content="Poems" />
```

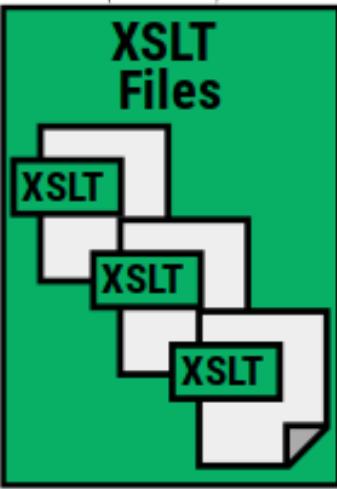
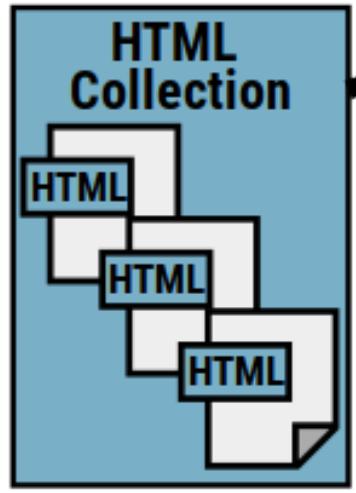


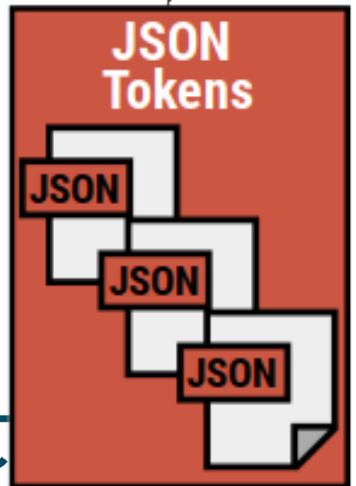
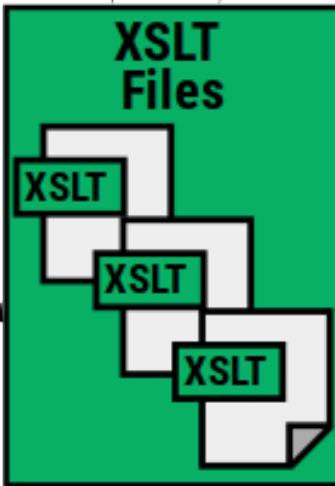
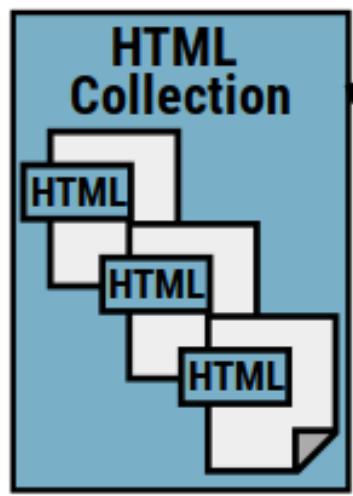
You create:

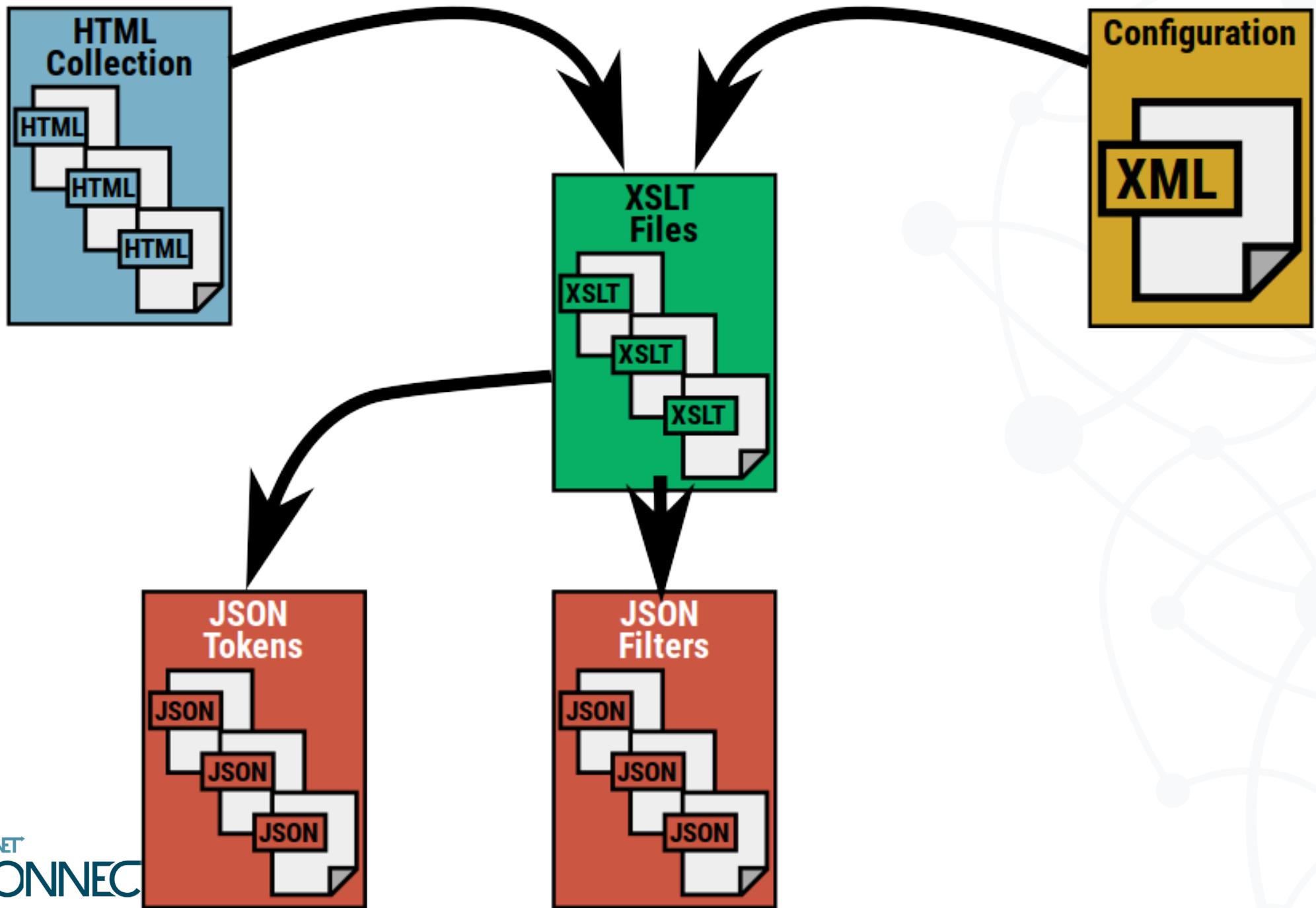


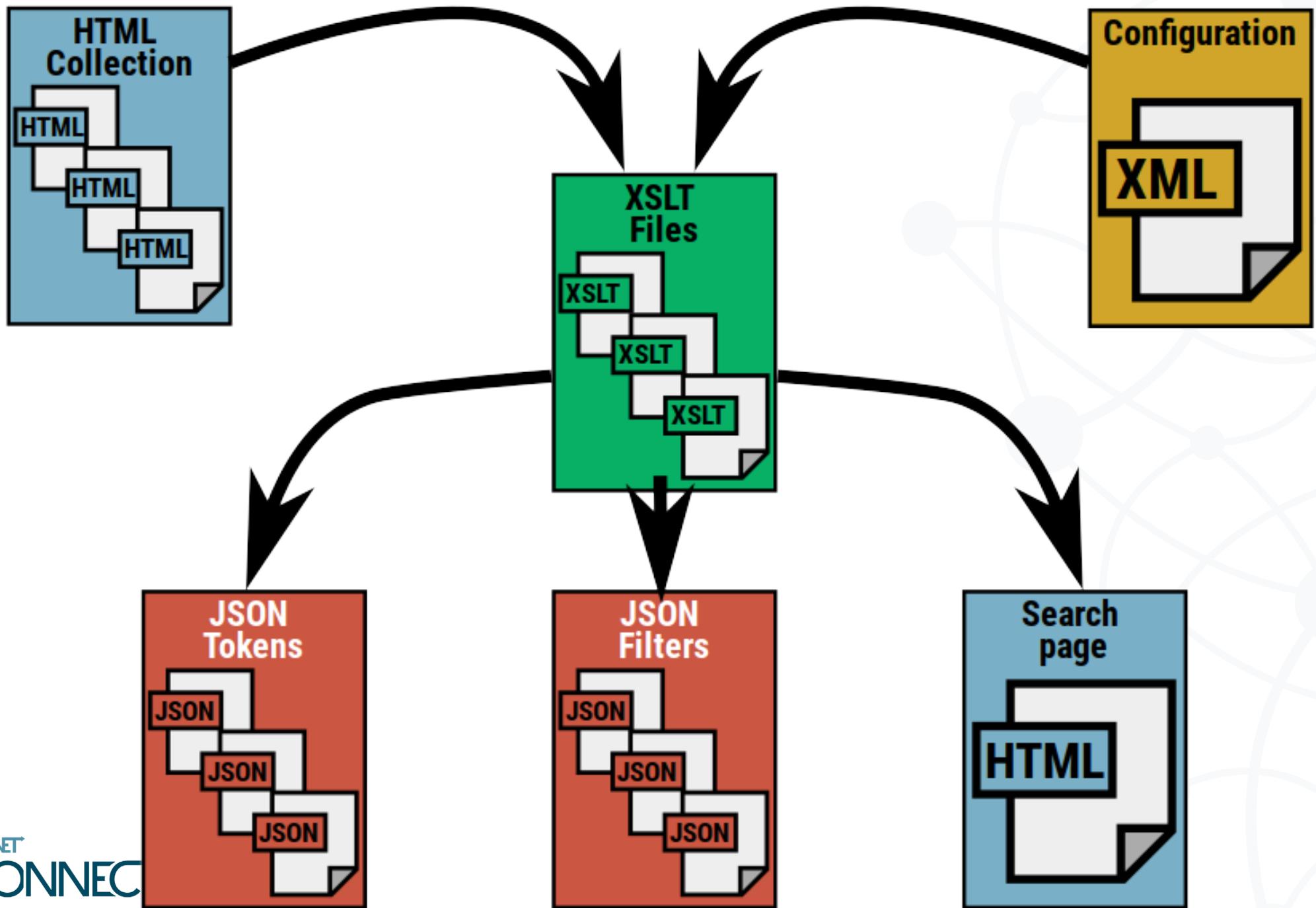
- A configuration file (XML) to specify the features and constraints for your search engine











JSON token file

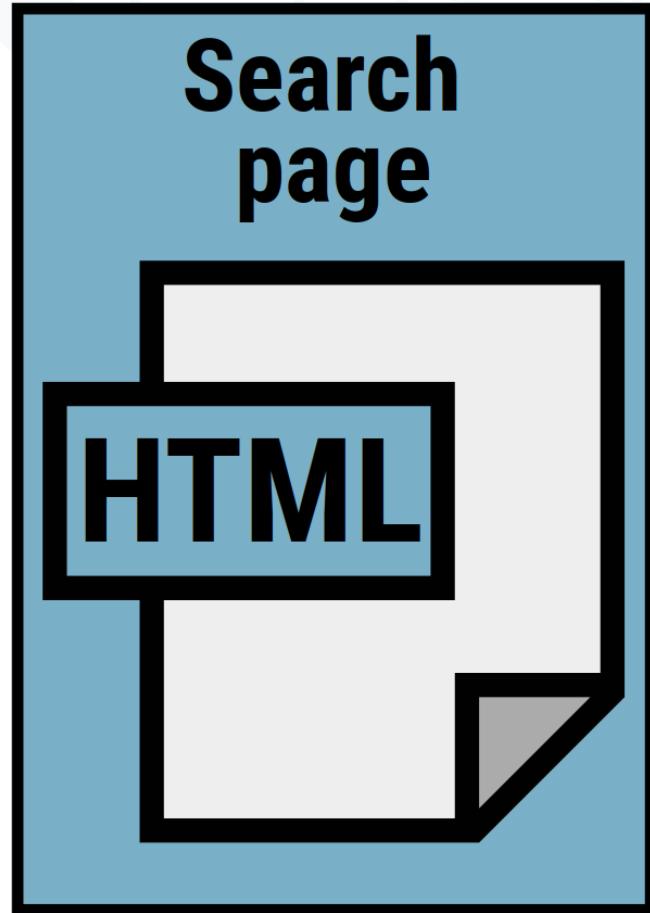
```
{"token": "unprofit",
 "instances": [
 {
 "docId": "pom_2025_ithe_old_man_of_hoy",
 "docUri": "poems/goodwords/1870/pom_2025_ithe_old_man_of_hoy.html",
 "score": 1,
 "contexts": [
 {
 "form": "unprofitably",
 "context": "...whole that day was spent <mark>unprofitably</mark>.",
 "weight": 1,
 "pos": 400
 }
 ],
 },
 {
 "docId": "pom_8733_john_and_joan_canto_ii",
 "docUri": "poems/blackwoods/1820/pom_8733_john_and_joan_canto_ii.html",
 "score": 1,
 "contexts": [
 {
 "form": "unprofitable",
 "context": "...too much ap- propriated unto <mark>unprofitable</mark> jocularities and facetiousness. Craving licence,...",
 "weight": 1,
 "pos": 164
 } ] } ] }
```

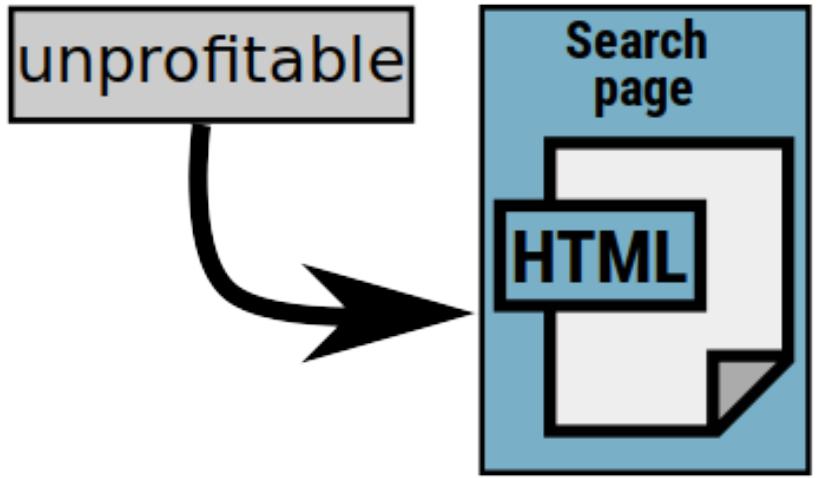
Filter JSON

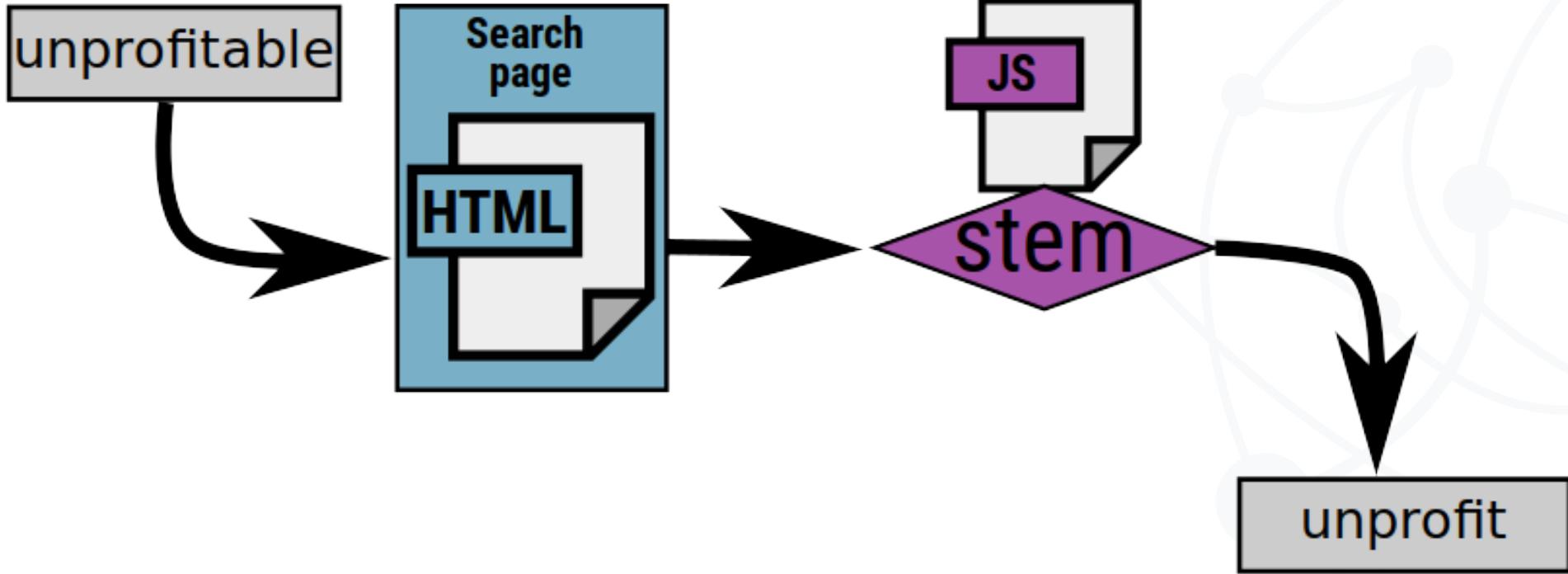
```
{  
  "filterId": "ssBool2",  
  "filterName": "Unsigned",  
  "ssBool2_1": {  
    "value": "true",  
    "docs": [  
      "poems/chambers_series/1867/pom_7594_the_husbands_request.html",  
      "poems/alltheyearround/1875/pom_4271_the_hourglass.html",  
      "poems/alltheyearround/1879/pom_4507_in_the_conservatory.html",  
      "poems/blackwoods/1843/pom_9963_jolly_father_joe.html",  
      "poems/blackwoods/1829/pom_10314_the_watchmans_lament.html"  
    ]  
  }  
}
```

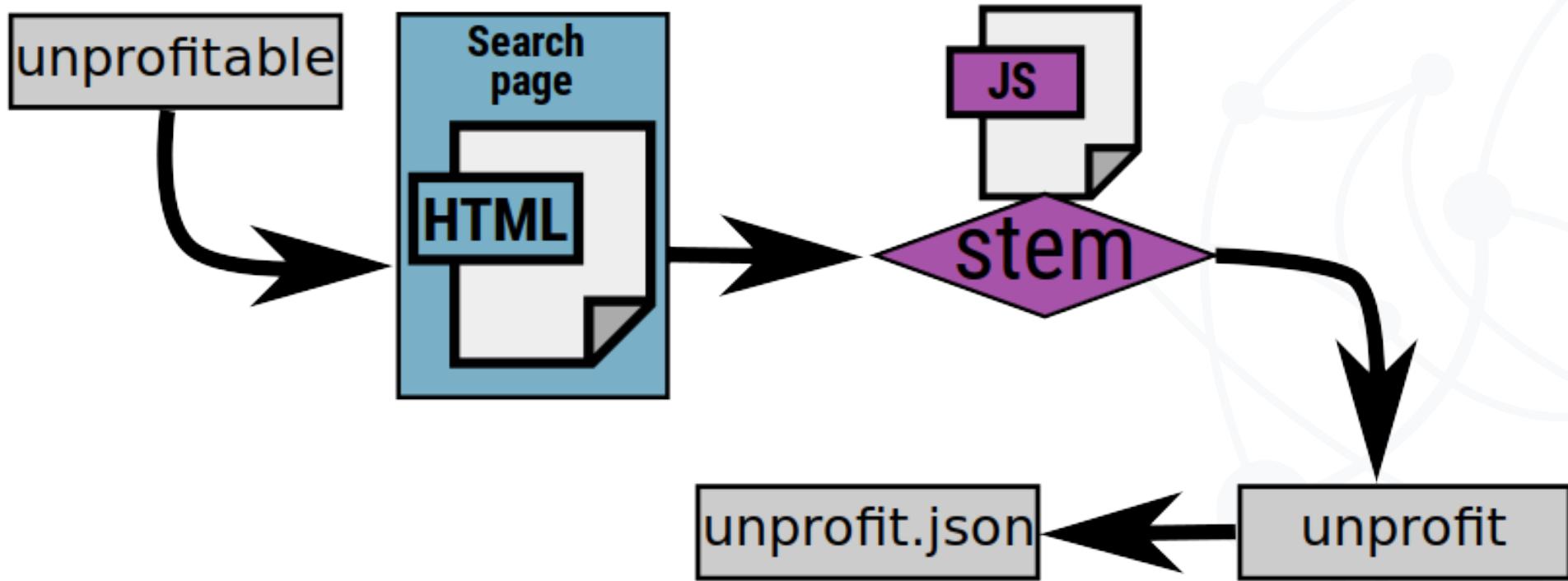
On the website...

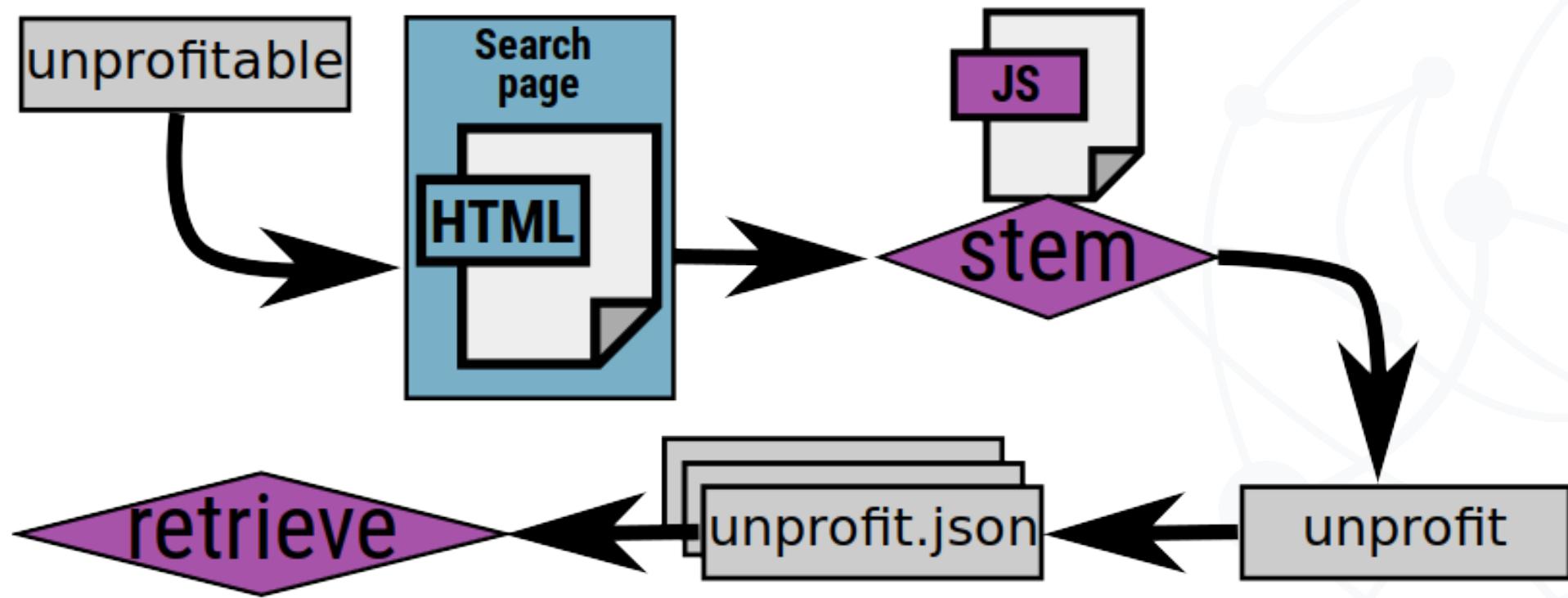
- ...of the thousands of JSON files...
- ...the search page JS retrieves only the ones it needs for the search you do:

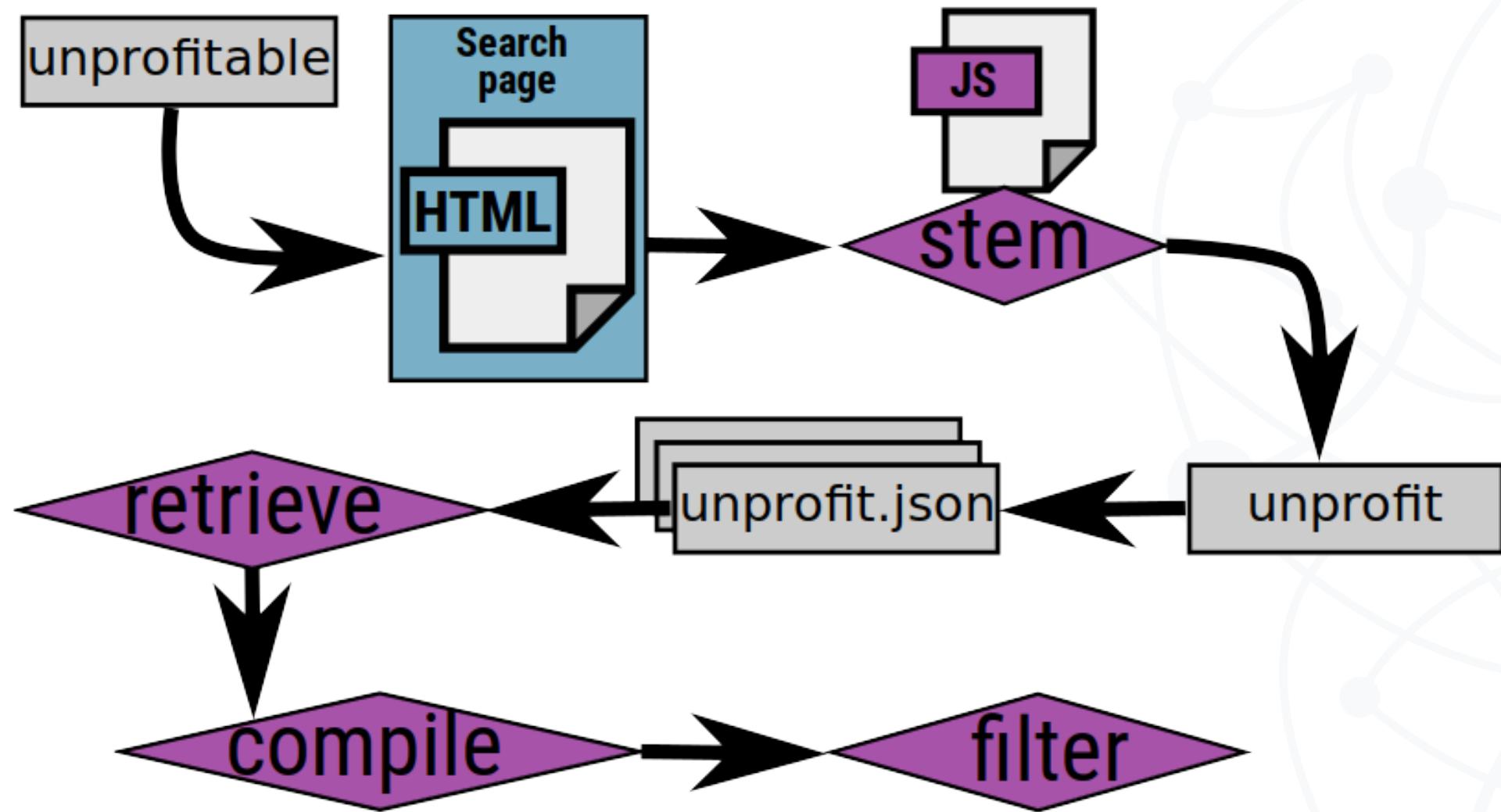


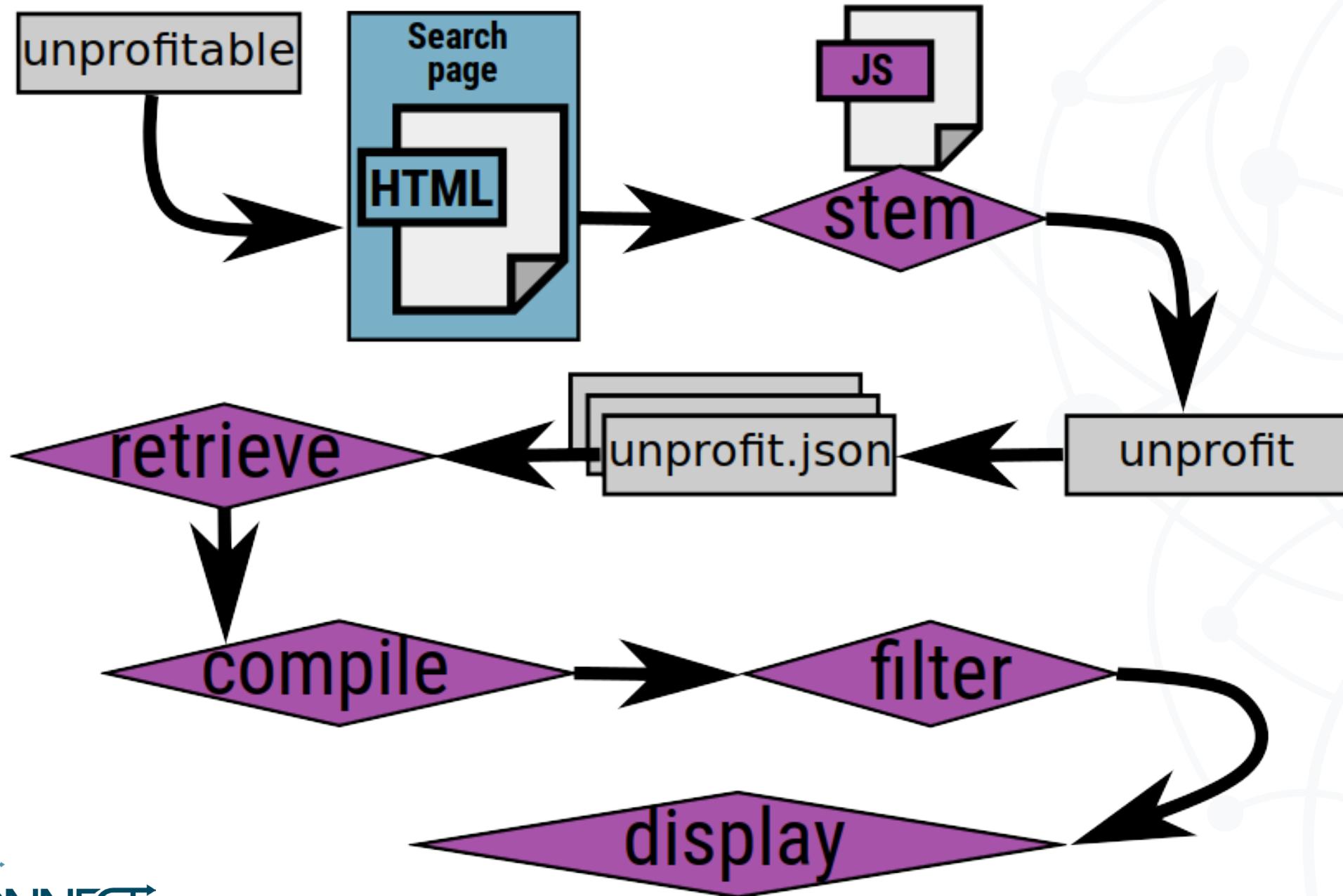












Documents found: 2

I.—The Old Man of Hoy Score: 1

- ...whole that day was spent unprofitably.

John and Joan, Canto II Score: 1

- ...too much ap- propriated unto unprofitable jocularities and facetiousness. Craving licence,...