Annotation-Based
Web Content Transcoding

Masahiro Hori
IBM Tokyo Research Lab.

Outline

■ **Annotation framework**
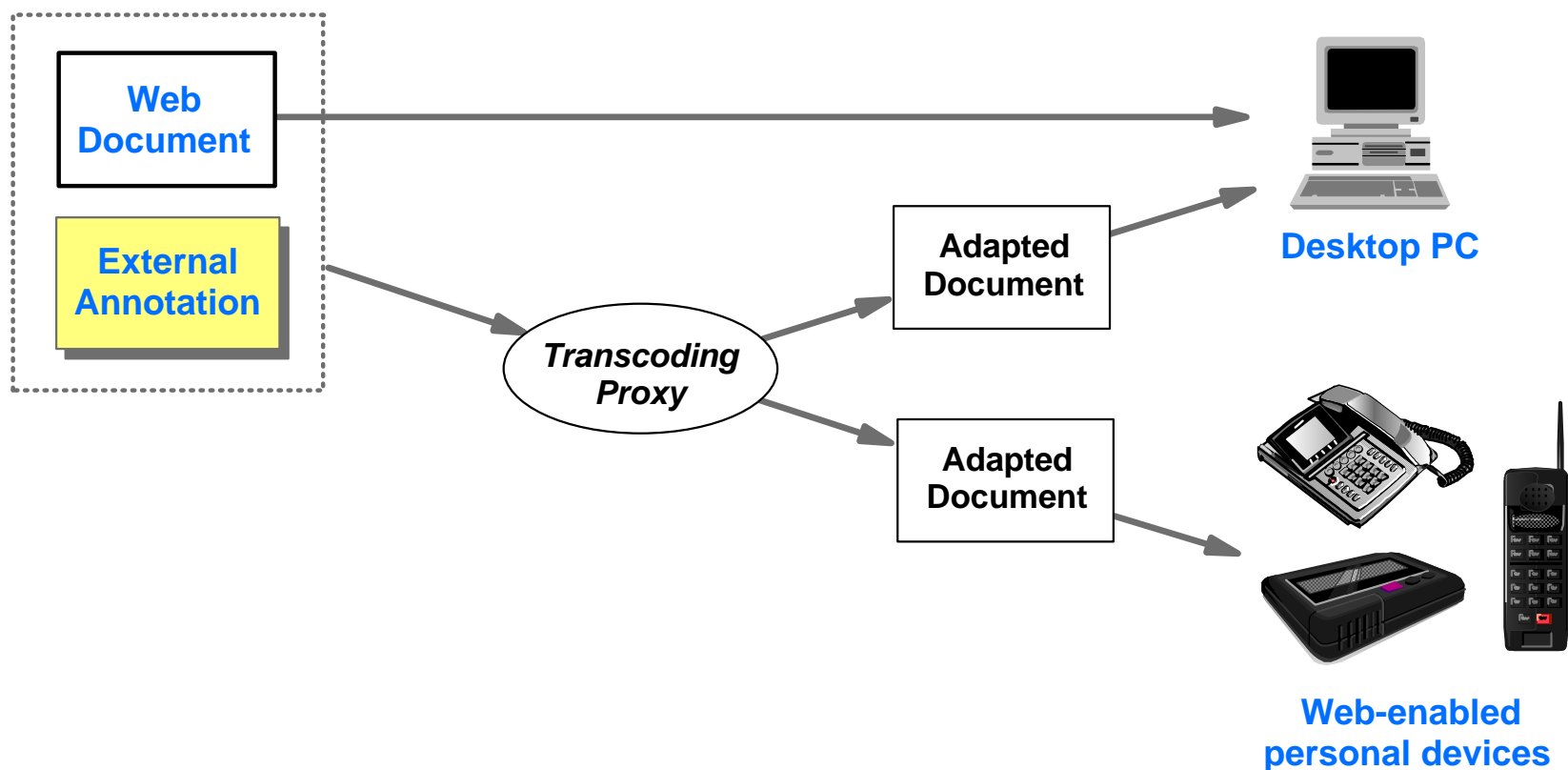
▸ Adaptation to Web-enabled personal-computing devices

▸ Vocabulary for adaptation hint

■ **Annotation-based transcoding**

▸ Demo of an annotation tool with a transcoding proxy

▸ Page splitting for small screen devices

Content Adaptation Scenario

- Augment Web documents with annotations, which provide hint for adapting them to a variety of client devices

```
Web
Document ─────────────────────────────► Desktop PC

External          Transcoding    Adapted
Annotation ──────► Proxy   ──────► Document ──► Desktop PC

                               ──► Adapted
                                   Document ──► Web-enabled
                                                personal devices
```

Web Document

External Annotation

Transcoding Proxy

Adapted Document

Adapted Document

Desktop PC

Web-enabled personal devices

External Annotation: Design Goals

- **Not introduce new tags/attributes into doc specification**
- **Allow annotating arbitrary part of target documents**

Web Document (HTML)

```
<HTML>
<HEAD>
 ...
</HEAD>
<BODY>
<H2>Turtle Tubo 999</H2>
<P> ... </P>
<OBJECT data="car1.mpg">

<P> ... </P>
...

</BODY>
</HTML>
```

External Annotation (RDF/XML)

```
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="..." >
 ...

<rdf:Description
    about="...#//child::P[2]">
  <pcd:priority value="0.2"/>
</rdf:Description>
 ...
<rdf:Description about= ...>
  <pcd:Group/>
</rdf:Description>

</rdf:RDF>
```

Annotate

XPath
XPointer

Vocabulary for Transcoding

- **Hint of content adaptation for personal computing devices**
  - ▶ Namespace prefix `"pcd"` (Personal Computing Devices)
- **Three types of annotation**
  - ▶ Selection criteria
    - `pcd:role, pcd:importance, pcd:fidelity, pcd:clientCapability, pcd:resourceRequirement`
  - ▶ Alternatives
    - `pcd:Alternatives, pcd:Replace, pcd:resourceToSubstitute`
  - ▶ Splitting hint
    - `pcd:Group`

See the W3C Note for further details
`http://www.w3.org/TR/annot/`

Example of Selection Criteria

- **Annotate a table element indicated by <TABLE> tag**
  - ▸ With a header role and the highest importance value
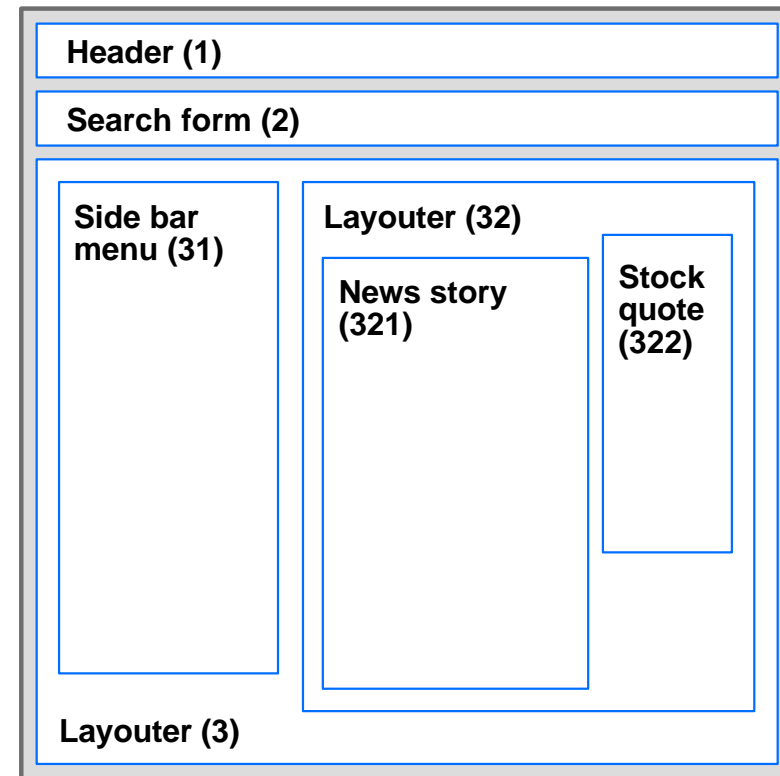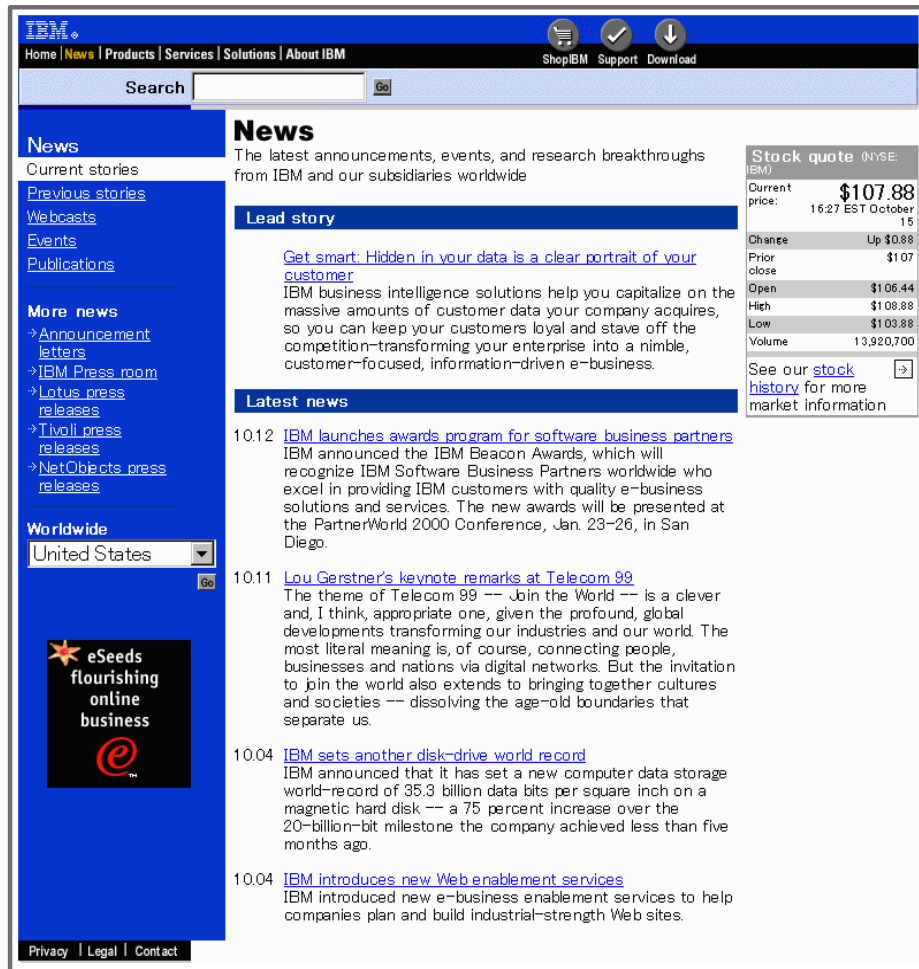
```
<rdf:Description
      about="http://foo.com/news.html#//child::TABLE[1]">
  <pcd:role value="header" />
  <pcd:importance value="1.0" />
</rdf:Description>
```

- **The annotated element is regarded as a common header for split-off pages**
  - ▸ Other roles: "footer", "auxiliary", "layout", "spacer" ...
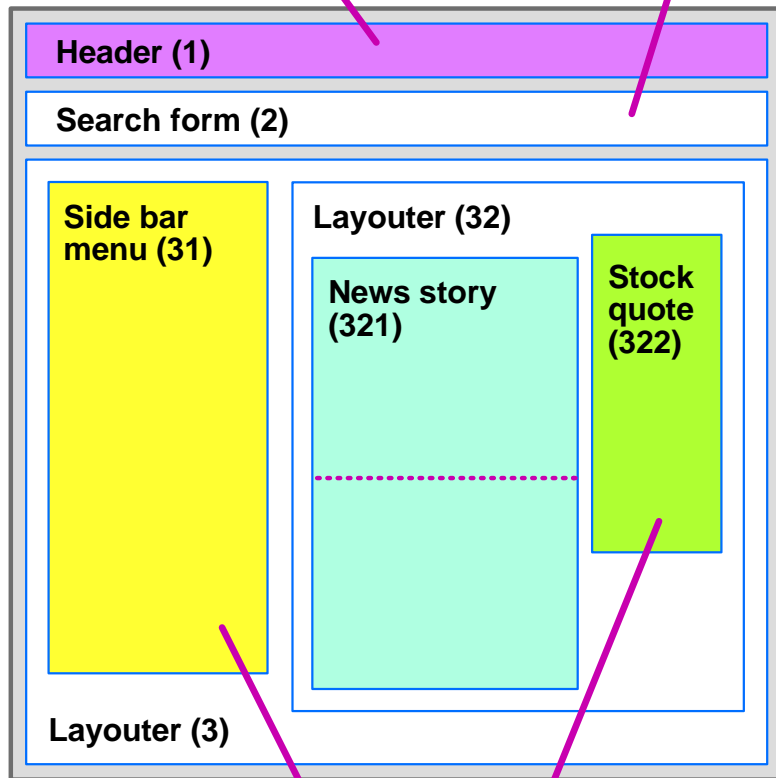- **Results depend on the interetation of a content-adaptation module**
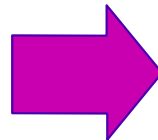
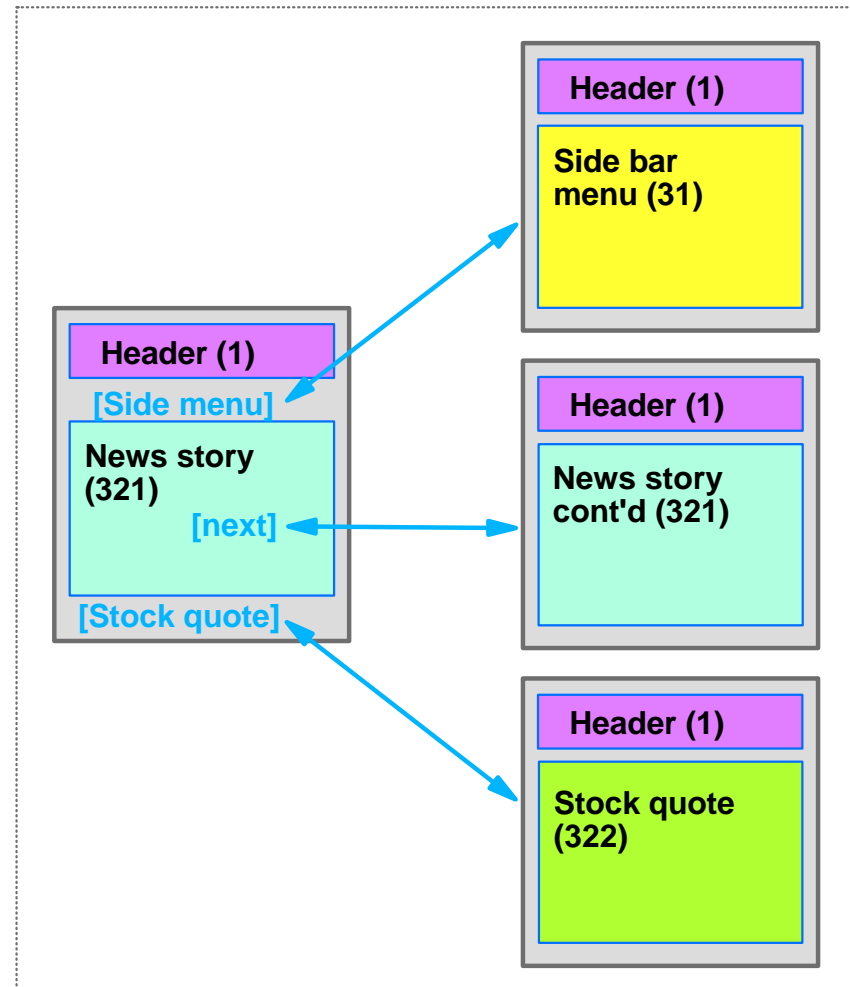# Layout of Real-Life News Page

`http://www.ibm.com/news/`

Header (1)

Search form (2)

Side bar menu (31)

Layouter (32)

News story (321)

Stock quote (322)

Layouter (3)

Fragmentation of News Page

role: header *(to be a common header for split-off pages)*

importance: -1.0 *(to be left out)*

Results of adaptation

Header (1)

Search form (2)

Side bar menu (31)

Layouter (32)

News story (321)

Stock quote (322)

Layouter (3)

role: auxiliary *(to be split off as a separate page)*

Header (1)

[Side menu]

News story (321)

[next]

[Stock quote]

Header (1)

Side bar menu (31)

Header (1)

News story cont'd (321)

Header (1)

Stock quote (322)
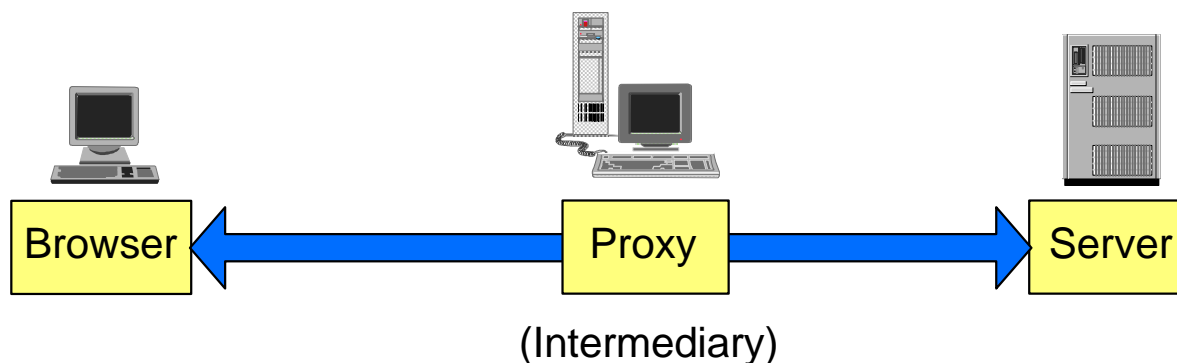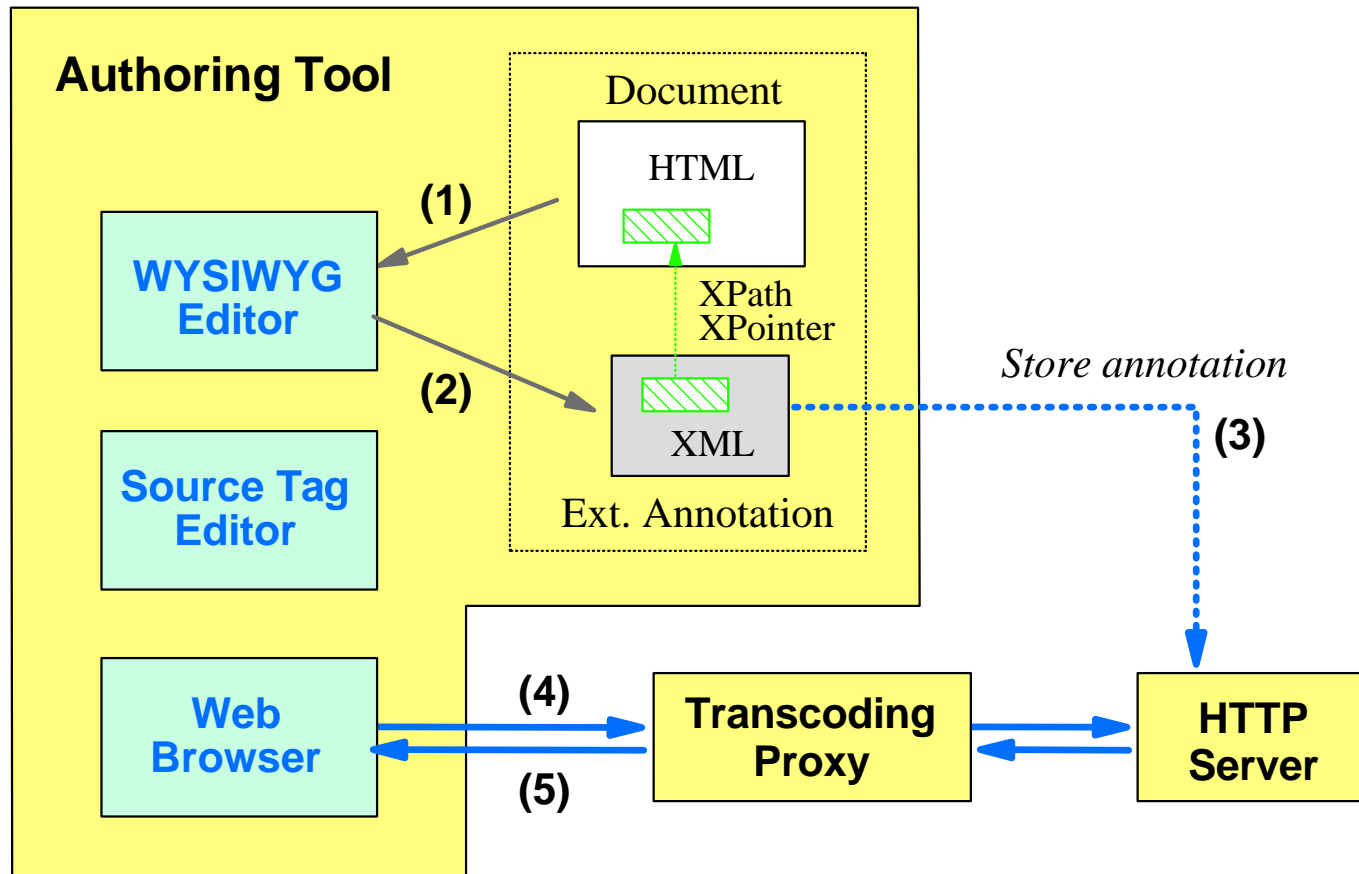
Transcoding Architecture

■ **Programmable proxy, or intermediary**
  ‣ Computational entities stay along the Web transaction path
  ‣ Operate on a document to be delivered over HTTP

■ **Usefulness for the annotation-based approach**
  ‣ Transform the contents with reference to external annotation
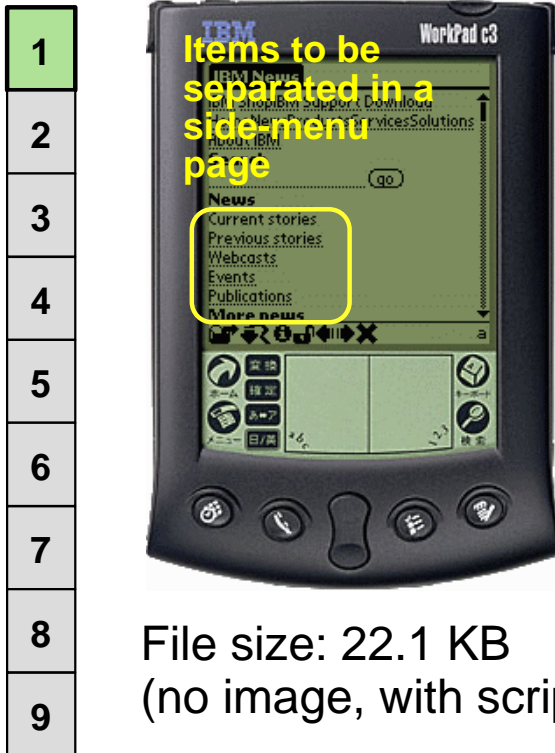 ‣ Without modification of content servers and Web borwsers

Browser		Proxy		Server

(Intermediary)

`http://www.ibm.com/software/websphere/transcoding`

Authoring-time Transcoding

Authoring Tool

Document

HTML

(1)

XPath
XPointer

WYSIWYG Editor

(2)

XML

Store annotation

(3)

Ext. Annotation

Source Tag Editor

Web Browser

(4)

Transcoding Proxy

HTTP Server

(5)

Comparison of Display on WorkPad

(a) Without page splitting

One ninth of the entire
document is on the display

File size: 22.1 KB
(no image, with script code)

(b) With page splitting

Half of the entire
document is on the display

File size: 6.8 KB
(no image, with script code)

Note: Rendered on an HTML browser (Palmscape).

Concluding Remarks

■ **Summary**

- ▸ The idea is to distinguish content with meta information
    - ● Facilitate articulation of implicit knowledge on the content use
    - ● Useful for existing (legacy) content
- ▸ Annotation provides <u>additional hints</u> for the purpose at hand
    - ● Need not be employed, if it can be done without ...
    - ● Play a role of mediating representation

■ **Future Works**

- ▸ Conduct further experiments by using real-life Web contents
    - ● To establish usability metrics for page fragmentation