

Using *Google Docs* to Collaborate on Papers and Study Books

Stijn Dekeyser (M&C)

Richard Watson (M&C)

Peter Sefton (DeC)

Contents

Richard

Motivation

Solutions

■ Public Domain

Email exchange / SharePoint

CVS or SVN with Latex or HTML

Google Docs (formerly *Writely.com*)

DEMO

■ USQ Solutions

GOOD / Epic

ICE

Stijn

Extending Google Docs

DEMO

Relation to research

Peter

Peter Sefton's insights on ICE and Google Docs

Motivation

- Typical Scenario
 - Multiple authors collaborate to produce a document
 - Collaborators may be geographically distant
 - We want tools to enable this to happen *simply and easily*
 - There should be little startup overhead (learning, client/server software installation etc)
 - For Mathematics and Computing:
 - Authors are using LaTeX/BibTeX, sometimes with CVS/SVN
-

Existing Solutions (1/4)

□ Email

- Documents developed locally
 - Authors work sequentially, not concurrently (synchronous)
 - Any format acceptable
 - onus on user to manage files and versions
-

Existing Solutions (2/4)

- CVS or SVN (Subversion)
 - Documents developed locally
 - Central repository of documents
 - Limited to ASCII-based formats (e.g. LaTeX, XML, HTML)
 - Provides 'almost' seamless concurrency
 - User intervention required if conflicting updates
 - chance for conflicts increases with time between commits
 - Syntax and semantics are non-trivial
 - Need to install CVS/SVN server and client software
 - Need host for server application & document repository
-

Existing Solutions (3/4)

- Google Docs
 - Provides 'almost' seamless concurrency
 - Conflicting update is rejected
 - However, conflicts are *very* rare
 - due to fine granularity and very frequent commits
 - Document held on Google server (can be downloaded) (mainly bad)
 - Very simple interface
 - Uses XHTML as underlying document model
 - See later in-depth discussion
-

Existing Solutions (4/4)

- USQ systems

- Good / Epic

- For Study Material, not research papers

- Use only within USQ

- ICE

- For any kind of document

- Open Source, GPL license

- See later discussion

Google Docs (1/2)

- Any (but only) gmail users can use Google docs
 - Three kinds of participants
 - owner (1)
 - collaborator (many; read/write)
 - viewer (many; read only)
 - Very simple word processor-like interface
 - Runs as client Javascript application on browser
 - Can "publish" to the web; accessible via a Google URL
 - Can save to local machine as HTML/OpenOffice/Word/PDF/RTF
 - Maintains revision history; revisions viewable
-

Google Docs (2/2)

- Performs automatic save (approx 30 sec)
 - significantly reduces chance of conflicts
- Docs are stored on Google server
 - Bad: trust and ownership
 - Good: automatic backup, no storage cost
- No support for formulas, bibliography, ...
- No tags for authors, abstract, keywords, ...
- Low control over layout
 - HTML, you can apply your own CSS style later
- No support for collaborative offline document update
 - Cannot merge a local document with a Google document

Problems

→ We propose solutions later

GOOD / Epic

- ❑ Used for study material, not research papers
 - ❑ Strongly structured, detailed content
 - ❑ Requires Epic editor on the Client
 - ❑ Can work off-line, but no rendering then
 - ❑ Locks document: only serial collaboration (synchronous)
 - ❑ Uses CVS underneath (transparent to user)
-

ICE

- Used for any document
- Uses HTML as underlying document model
- Can be used with MS Word and OO Writer
- Requires **OO.org**, **Web Server**, **SVN Client** on Client
- Currently difficult to set-up and use
- Can work off-line, versions merged thru SVN
- Can export to PDF, DOC, HTML
- No support for formulas, bibliography, ...
- No tags for authors, abstract, keywords, ...
- infrequent commits increases onus on user to resolve conflicts
- Low control over layout

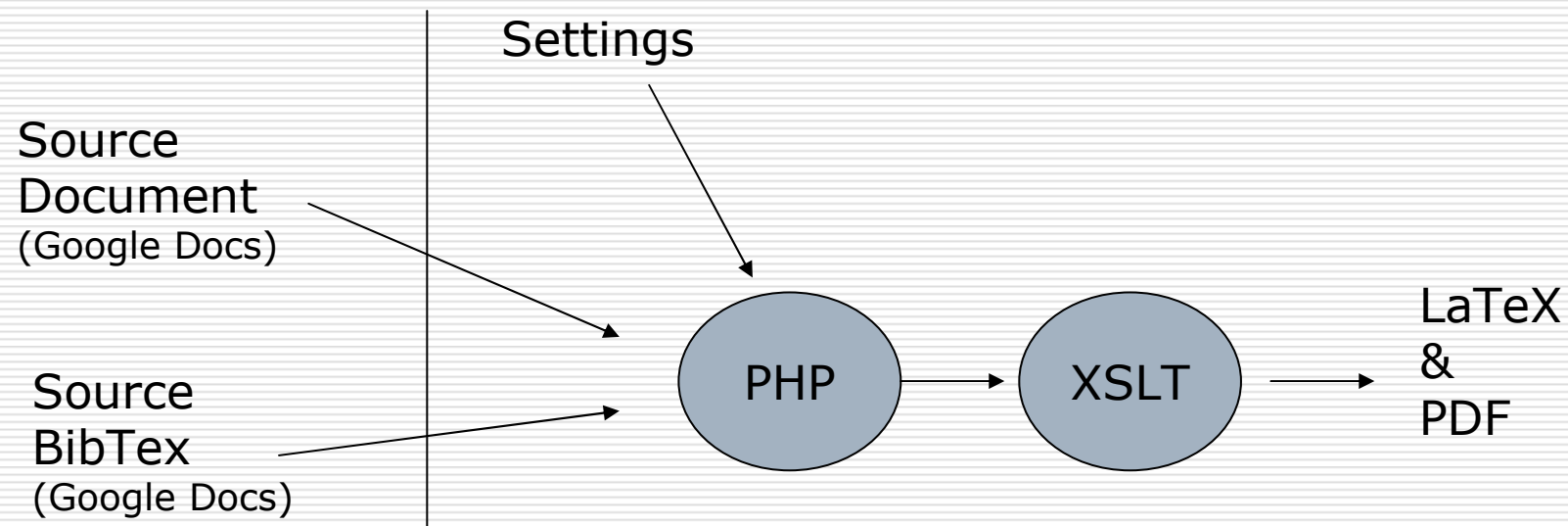
Peter Sefton will present his insights later

Subjective Comparison

- GOOD:
 - + use (for study material only) if happy with Epic
 - - not good for concurrent collaboration
 - ICE:
 - + good for on-line AND off-line collaboration
 - - not fully ready yet, and lots of software on client
 - - infrequent commits increases risk of conflicts
 - - weak control on layout
 - Google Docs:
 - + excellent for live, ad hoc collaboration
 - - not good for off-line collaboration
 - - weak control on layout
-

Extending Google Docs (1/4)

□ For Research Papers:



Extending Google Docs (2/4)

- Both problems with Google Docs can be solved with 'simple' extensions
 - Our first extension solves layout, bibliography, formulas etc
 - Simple *proof-of-concept* implemented in < 4 hours
 - Another extension we thought of could solve off-line collab problem
 - Not implemented yet
-

Extending Google Docs (3/4)

- Issues:
 - Safety problem (LaTeX is Turing Complete)
 - See next slide
 - Use of packages not present on the server
 - More advanced bibliography solution may be needed
 - Needs a more user-friendly form for non-Latex people
 - Our advise:
 - Use Google Docs only as a flat text editor
 - Don't use the tables or lists; use LaTeX markup
 - If you don't know LaTeX:
 - you may use tables, lists etc, but problems may occur in the translation
-

Extending Google Docs (4/4)

- ❑ LaTeX is a Turing Complete language
 - ❑ Run on the web server as user www
 - ❑ Form text boxes allow any LaTeX code
 - ❑ Three possible solutions:
 - Running as a local user, not a web app
 - Put on a virtual web server
 - Replace free-input by specific input
-

Relation to Path Locks research

- Path Locks for XML concurrency control is one of my research areas
 - Comparison to Google Docs:
 - + Path Locks ensure serializability (no conflicts ever need user attention)
 - + Path Locks are general: for any XML doc (including graphics)
 - - Path Locks are less efficient
 - - Granularity of Path Locks is per XML element; much coarser than Google Docs
-

Peter Sefton's Insights

ICE Update

- Now has installers
 - ICT can install for USQ users – we have SVN servers running for you
 - ICE handles books / IMS course packages
 - Bibliographies and citations using Word + Endnote
 - Maths via MathType (LaTeX soon)
 - A server-based version coming in 2007
 - Integrated presentation software
-

Google Docs issues

- ❑ Won't import and export word processing documents
 - ❑ Produces bad, broken HTML even in simple cases
 - ❑ How would one export and import courseware packages?
 - ❑ My observations:
<http://del.icio.us/ptsefton/writely>
-

Stuart's list:

1. My point

a. Ordered entry

b. Ordered entry

c. Ordered entry

1. My point

• Unordered entry

• Unordered entry

```
Stuart's list:<br/>
<ol>
  <li>
    My point</li>
</ol>
<ol style="MARGIN-LEFT: 40px" type="a">
  <li>
    Ordered entry</li>
  <li>
    Ordered entry</li>
  <li>
    Ordered entry</li>
</ol>
<ol>
  <li>
    My point</li>
</ol>
<ul style="MARGIN-LEFT: 40px">
  <li>
    Unordered entry</li>
  <li>
    Unordered entry</li>
</ul>
```

Use Styles!

- A well-designed style sheet can allow interoperability between web and word processing documents
 - Styles can imply structure
 - h1 ... h5
 - List 1 ... List 5
 - Even without styles the Google Docs formatter could be much smarter
-

End

Questions?

Comparison

	GOOD	ICE	Google Docs
Schema	GOOD DTD	XHTML	XHTML
Editor	Epic (Windows)	MS Word and OO Writer	Browser (XHTML editor)
Synchronization	CVS checkout 1 ed at a time	SVN Multiple eds	Live synch Multiple eds
Software	Epic	Web Server Open Office	Browser
Location of documents	Central USQ store	Distributed on clients	On Google server
Formulas	LaTeX support	?	-- (use LaTeX)

Comparison

	GOOD	ICE	Google Docs
Export	XML HTML PDF	HTML PDF (Word & OO)	XHTML MS Word OO Writer RTF PDF
