

SISRS: The Superimposed Scholarly Review System

Sudarshan Murthy, David Maier

Department of Computer Science, Portland State University

PO Box 751

Portland, OR 97207-0751

+1-503-725-4036

{smurthy, maier}@cs.pdx.edu

Created on: November 05, 2004; last modified on: February 24, 2007.

ABSTRACT

Reviewing scholarly work in electronic form (for example, in PDF) is becoming increasingly commonplace. Applications such as Adobe® Acrobat® and Microsoft® Word facilitate annotating, an important part of the review process, but they have some drawbacks. For example, they store annotation in the reviewed document, and they provide poor query capability. The *Superimposed Scholarly Review System* (SISRS) addresses these drawbacks. It *superimposes* reviews and leaves source documents unchanged. It uses a common user interface over heterogeneous document types and it integrates into host applications such as Acrobat and Word. It allows aggregation of comments by multiple reviewers, and querying of comments and commented regions together. It even allows multiple simultaneous organizations for a review. In addition to demonstrating an alternative means of scholarly review, SISRS demonstrates some of the benefits of superimposed information and queries over bi-level information. SISRS is built using SPARCE, a middleware architecture for superimposed information management.

1. INTRODUCTION

We see three significant activities in the traditional scholarly review of a paper: reviewers independently review the paper, each reviewer writes a report and submits it to a central location, and the author receives a collection of reviews and uses it to improve the paper. Applications such as Adobe Acrobat and Microsoft (MS) Word facilitate commenting over documents (support for the first activity), but they have some drawbacks. For example, they embed comments in the source document, thus altering it. They limit the structure of a comment, and generally do not allow expression of inter-comment and inter-document relationships. They have limited navigation and query capability. Also, they lack support for report generation. Current electronic systems (such as conference review systems) typically aid in parts of the second and third activity. They help in submitting individual review reports and in creating the collection of reviews for a paper.

The *Superimposed Scholarly Review System* (SISRS; pronounced *scissors*), improves assistance for all three activities. It assists individual reviewers in reviewing papers in electronic form (such as PDF documents) and in preparing review reports. It also aids facilitators' and authors' understanding of review comments. It complements the electronic systems currently in use.

Some of the capabilities of SISRS are:

- Create *superimposed* comments over heterogeneous document types such as PDF and Word in their *native applications*, but with a common user interface.
- Customize comment structures (we will show two structures).
- Aggregate reviews by multiple reviewers.
- Navigate easily from a comment to commented region.
- Relate comments, commented regions, and outside information.
- Generate review reports, and create alternative organizations of comments and commented regions.
- Query over comments, commented regions, and relationships.

We describe the key features of SISRS using some actual review comments, shown in Figure 1. The comments were recorded in a central repository, and communicated to authors of the paper by email (SISRS was not involved).

Section 5, 2nd paragraph, last line:
"... but it ha ..." typo.

Page 2:
You gave a definition for "item" and a definition for "group".
What is the definition of a "document"?

What happens if the underlying documents change? Do you assume only static documents?

Figure 1: Example review comments

The first line of the first comment specifies a region of the paper, the quoted text in the second line provides some context (partial excerpt), and the last word 'typo' succinctly states a problem with the identified region. The first line of the second comment specifies a region of the paper. The second line states that the authors omitted the definition of a concept. The context for the comment is only indirectly specified. The last comment contains some questions, but no location or context information.

In general, a review comment has three parts: address of the region of a document to which the comment applies, some context information (such as an excerpt), and the text of the comment (such as a question). However, as in the comments of Figure 1, the specificity of the address of the commented region and the precision of context information vary widely. For example, the context information in the first comment is apparently *precise enough*, however the context information is missing in the third

comment, possibly because it is quite difficult to obtain it (or specify it). The address is too broad in the second comment (Page 2), likely because it is tedious (or unnecessary) to specify the locations of definitions of various concepts (possibly spread over Page 2).

SISRS automatically generates address (both machine-friendly and human-friendly) and context information for comments. A reviewer simply reads a document in its native application (for example, PDF in Adobe Acrobat or HTML in a browser), selects a region of the document and comments on it. Figure 2 shows SISRS being used, from within MS Word, to comment over a region of a Word document. Section 3 provides details.

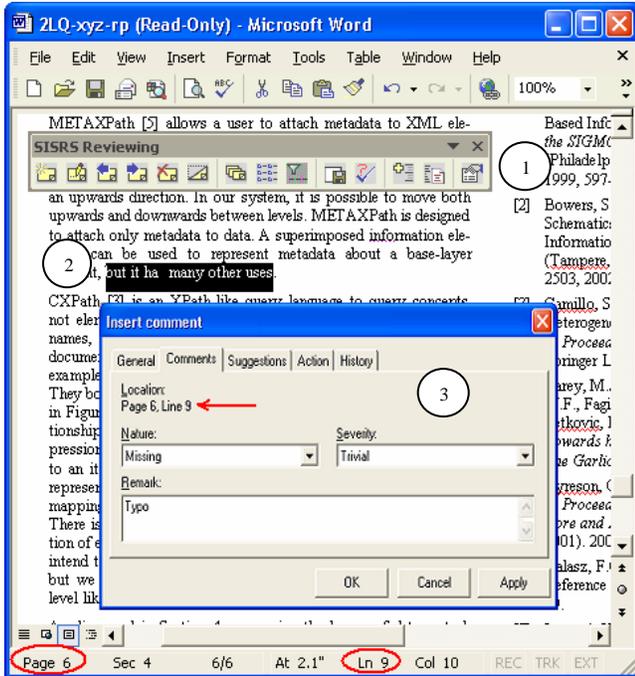


Figure 2: Creating a review comment with SISRS

SISRS allows a reviewer to specify relationships between comments and between regions of different documents. For example, a reviewer may highlight an inconsistency between parts of a document, or a reviewer may compare a commented region with information in a cited paper.

SISRS supports navigation and queries over comments, commented regions, and relationships (details in Section 3). In addition to filtering, grouping, and aggregating, queries can be used to generate review reports, and to create alternative organizations of reviews. Some possible queries and example uses are:

Q1: List comments, of all reviewers, whose severity is "major" (assumes the comment structure includes severity assignment)

- Filter; allows a committee to get an overview.

Q2: List comments, of all reviewers, in document order (by location of commented regions)

- Grouping; generates a review report, creates an alternative view.

Q3: Show the document page with most comments

- Aggregation; identifies a potential problem area.

Q4: List headings of sections that contain comments whose severity is major

- Projection and navigation; allows an author to prioritize work.

Q5: List documents a reviewer has cited in the report

- Navigation to external documents; exploits relationships, allows an author to look at external literature referred to by a reviewer.

We see several applications for SISRS. For example, it can be used in document inspection and source code inspection, when conducting background research, and to annotate medical records.

2. SUPERIMPOSED INFORMATION

Superimposed information is data placed over existing information sources to help select, access, organize, connect, and reuse information elements in those sources [2]. Existing information sources reside in the *base layer*, and data placed over one or more base sources resides in the *superimposed layer* (see Figure 3). Documents that are reviewed (word-processor documents, web pages, and so on) are examples of base documents. A review comment in SISRS is an example of superimposed information (because it is stored separately from the commented region, and it maintains a link to the region). An application that manipulates base information is called a *base application*; an application that manipulates superimposed information is called a *superimposed application*. MS Word is a base application, SISRS is a superimposed application.

A superimposed information element (such as a comment) refers to a base information element (such as a commented region in a MS Word document) using an abstraction called a *mark*. Figure 3 shows the superimposed layer using marks to address base elements. Several implementations of the mark abstraction exist, typically one per base type. In order to address a selection in a base document, a mark implementation supports an addressing scheme appropriate for that base type. For example, an implementation for PDF documents may use page number and index of the first and last words in a text selection, whereas an implementation for MS Word documents may use starting and ending character indexes. All mark implementations provide a common interface to address base information, regardless of the base types or access protocols they support. A superimposed application can work uniformly with any base type by virtue of this common interface.

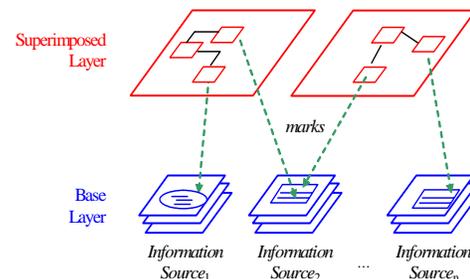


Figure 3: Layers of information

Superimposed applications may occasionally need to incorporate the content of base-layer elements in the superimposed layer. For example, a review comment may include the text of the commented region. The content of a base-layer element is called an

excerpt. An excerpt can be of various types. For example, it may be text or an image.

In addition to excerpts, superimposed applications may use other information related to base-layer elements. For example, a review report may group comments by sections of the source document. To do so, SISRS retrieves the section heading (assuming one exists) of each commented region. Information concerning a base-layer element, retrieved from the base layer, is called its *context*. Section heading and presentation information such as font name might be included in the context of a mark. Each such piece of information is a *context element*, and context is a collection of context elements.

SISRS is built using the *Superimposed Pluggable Architecture for Contexts and Excerpts* [2] (SPARCE), our middleware architecture for superimposed information management. Superimposed applications use SPARCE to create marks and access context. SPARCE stores details of each mark, such as the address of the container and the selection inside it, as an XML file in a marks repository. It uses a mediator called *context agent* to retrieve context information for a mark from an appropriate base application.

3. SISRS DETAILS

SPARCE includes mark and context-agent implementations for PDF, MS Office documents, HTML, XML, and many audio/video file formats. SISRS reuses the implementations to identify and query selections in these types of documents. Our development toolset includes MS Visual C++, MS Visual Basic, and MS XML 4.0 SDK.

SISRS is integrated into Adobe Acrobat, MS Office applications such as Word and Excel, and MS Internet Explorer. (Our examples use MS Word only for illustration.) A user reviews a document or reads reviews of a document in an application native to the type of that document (Acrobat for PDF, MS Internet Explorer® for HTML, etc.). SISRS is only an add-in to these applications.

3.1 Information Model

SISRS represent reviews (superimposed information) as XML. Figure 4 shows a partial XML representation of a review (corresponding to the first comment of Figure 1). The element Comment represents a comment. The sub-element Related represents the relationships of a comment. SISRS does not itself maintain mark and context information (such as the address of a commented region and the excerpt of the region for a comment). Instead, it maintains the ID of the mark for the commented region, and retrieves the needed information from SPARCE.

```

- <Review DocumentID="2LQ-xyz-rp">
- <Comment ReviewerName="sms">
  <Mark ID="WordMark2004Jul31191750">...</Mark>
  <Nature>Missing</Nature>
  <Severity>Trivial</Severity>
  <Remark>Typo</Remark>
- <Related>
  <Mark ID="PDFMark2004Jul31192550">...</Mark>
</Related>
</Comment>
</Review>

```

Figure 4: XML representation of a review

The structure of a comment is customizable. SISRS defines an interface for plugging in different comment structures and user interface (UI). The custom comment implementation only needs to work with the element Comment, and provide a UI for its part of the data. SISRS takes care of integration into host applications and interaction with SPARCE. SISRS includes a default comment implementation that manages the elements Nature, Severity, and Remark inside the element Comment in Figure 4. It also provides the UI labeled 3 in Figure 2. The default comment structure can be replaced easily, without recompiling SISRS. Also, the replacement may use any structure (as long as it is modeled in XML and can be embedded inside the Comment element). For example, a custom implementation could use a structure based on the IEEE 1044 standard [1] to record defects in a program's source code.

Figure 5 shows the SISRS information model in an ER diagram. The entity type CommentDetails is supported by the comment implementation. The relationship type References represents the mark to the commented region. The relationship type Related to represents the marks to related sections of the subject document or other documents.

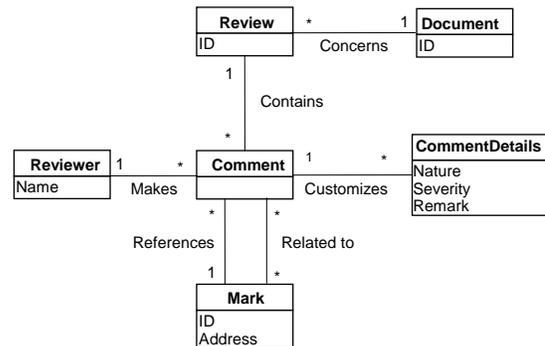


Figure 5: SISRS information model

3.2 Reviewing

A reviewer reviews a document using an application in which SISRS is integrated. Figure 2 shows a comment being created in MS Word (corresponding to the first comment of Figure 1). The numeric annotations in the figure call out three key parts. Part 1 is the 'SISRS Reviewing' command bar integrated into MS Office applications. This bar facilitates operations such as creating and editing a comment, and querying. Part 2 is the commented region. Currently, a reviewer has selected this region and is creating a new comment over that region. Part 3 is a tabbed dialog with details of the comment. The dialog shows the location of the commented region generated automatically (*Page 6, Line 9*, pointed to by an arrow), and verified on Word's status bar (circled). The dialog also shows the values the reviewer has entered for the attributes Nature (*Missing*), Severity (*Trivial*), and Remark (*Typo*).

3.3 Querying

Consider query Q4 mentioned in Section 1 (*List headings of sections that contain comments whose severity is major*). Answering this query requires retrieving the section heading of each commented region, but only for comments that meet a certain criteria. In general, queries in a superimposed information setting are

expressed against both superimposed information and base information. Consequently, they are called *bi-level queries* [3].

SISRS supports bi-level queries over comments, commented regions, and relationships. Queries may be expressed in XPath, XSLT, and XQuery. For ease of use, SISRS includes some predefined queries for filtering, grouping, and aggregation. For example, queries Q1 and Q2 are predefined. Executing query Q1 requires a user to supply a value for the severity parameter.

Here are two bi-level queries expressed as XPath expressions:

Q1: List comments whose severity is "major"
//Comment[Severity='Major']

Q5: List documents a reviewer has cited in the report
//Comment/Related/Mark/Container/Location

Query Q5 refers to elements not shown in the XML representation in Figure 4. Briefly, the element *Mark* contains details such as the location of the document (*container*), and the address within the document where the mark is made.

A user may execute ad-hoc queries, and optionally install them for future reuse. Because all users may not be proficient in writing XML queries, we expect queries to be developed as a part of the SISRS adoption process. For example, a conference program committee, with help from its IT support personnel, might build in appropriate queries with its distribution of SISRS. Alternatively, it might provide a web-based query service to avoid distributing custom queries.

3.4 Review reports

A review report is a transformation of one or more reviews. SISRS generates a query report using a predefined bi-level query (such as *Q2: List comments in document order*). Figure 6 shows part of a review report generated as an HTML document. We used an XSLT style sheet to generate this HTML. The report is currently opened in an HTML browser, not in SISRS. That is, review reports SISRS generates can be used independent of SISRS.

The comments in this report correspond to the comments of Figure 1, but they are displayed in document order of the commented regions. For each comment, the report includes the address of the commented region (such as *Page 6, Line 9*), the excerpt of the commented region, the remark text, and severity and nature of the problem identified. The heading 'Excerpt' also includes a link. Activating this link opens the reviewed document in the appropriate base application and makes the commented region visible. For example, activating this link for the last comment in the report opens the MS Word document of Figure 2 and highlights the text "but it ha many other uses."

Using a review report is quite easy. For example, a reviewer may simply paste the text of a report in to a review system (similar to

what is done in current systems). Alternatively, a reviewer may submit the HTML report. With appropriate software (included with SISRS), users may activate links in an HTML report to visit commented regions in reviewed documents.

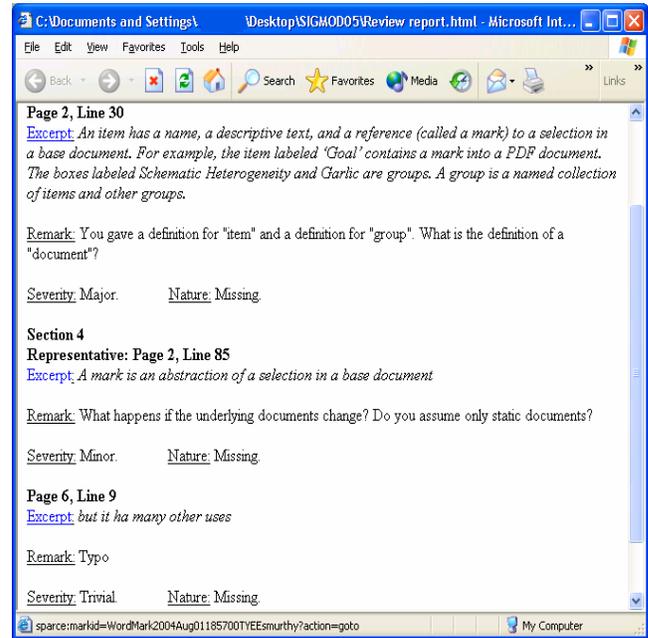


Figure 6: A review report

4. ACKNOWLEDGMENTS

We thank Lois Delcambre for her review. We thank the erstwhile Sunlet Software Systems, Inc. for allowing us to reuse and adapt some of their components. The National Science Foundation grants IIS-0086002 and IIS-9817492 supported this work.

5. REFERENCES

- [1] IEEE, Inc. IEEE Standard Classification for Software Anomalies (IEEE Standard 1044). IEEE 345 East 47th Street New York, NY 10017-2394. 1993.
- [2] Murthy, S., Maier, D., Delcambre, L., and Bowers, S. Putting Integrated Information in Context: Superimposing Conceptual Models with SPARCE. In *Proceedings of the First Asia-Pacific Conference of Conceptual Modeling* (Dunedin, New Zealand, Jan. 22, 2004). 71-80.
- [3] Murthy, S., Maier, D., and Delcambre, L. Querying Bi-level Information. In *Informal Proceedings of WebDB '04* (Paris, France, June 17-18, 2004). 7-1