Integrating Digital Libraries Through Linking
(NSDL Digital Library Service Integration Project)
Xin Chen*, Dong-ho Kim**, Nkechi Nnadi*, Michael Bieber*, Il Im* and Yi-Fang Wu*

*New Jersey Institute of Technology (NJIT)
University Heights, Newark, NJ 07102, USA
**Rutgers University Graduate School of Management
University Heights, Newark, NJ 07102, USA

KEYWORDS
digital library integration, service integration, automatic link generation, collaborative filtering, lexical analysis

INTRODUCTION AND MOTIVATION
This research provides a general method for integrating digital libraries and other Web systems through linking the interrelated elements and functions. While our approach is a general one, thus far we have applied it primarily to digital libraries.

The Digital Library Integration Infrastructure (DLII) automatically generates links for digital library collections to related collections and documents. Collections are libraries of computerized documents. Services include searching, providing annotations and peer review. Figure 1 presents an example from our current prototype, which can be accessed from our Web site. In addition to NASA’s National Space Science Data Center (NSSDC) and the Arizona Document Summarizer, we currently have three preliminary, partial integrations of digital library systems within the National Science Digital Library [http://www.nsdll.org]. These include the AskNSDL “ask an expert” service, the Atmospheric Visualization Collection, and the Earth Science Picture of the Day system, as well as MapQuest, amazon.com and the NJIT library.

DLII supplements collections by linking them automatically to relevant services and related collections. DLII supplements services by automatically giving relevant objects in collections (and other services) direct access to these services. Users see a totally integrated environment, using their system just as before. However, they see additional link anchors, and when clicking on one, DLII will present a list of supplemental links. The next release of DLII will filter and rank order this set of generated links to user preferences and tasks.

The DLII infrastructure provides a systematic approach for integrating digital library systems, and by extension, any other information system with a Web interface. Systems generally require no changes to integrate with DLII.

DLII INFRASTRUCTURE AND INTEGRATION
Figure 2 presents the DLII integration infrastructure. To integrate a system that requests information (where DLII provides links), an analyst must write an integrator. To integrate a system that provides services (to or through which DLII-provided links lead), an analyst must declare relationship rules and/or register glossaries/thesauri.

1. Develop an Integrator: An integrator sends DLII a preliminary list of elements eligible for linking within its system’s documents and screens. DLII uses “structural analysis” (relationship rules) to generate links for each element.

Several integrator approaches exist. External approaches, such as wrappers and content analysis operate solely on the collection or service’s output, and therefore require no changes to the system itself. This is especially useful for retrofitting DLII support to an existing system.

Internal approaches can generate element information as the screen or document is being produced, and either embed the information within it (e.g., using XML) or provide this information separately.
perhaps through an application programming interface (API).

(2) Declare Relationship Rules: Relationship rules specify the “structural relationships” for automatically generating links for recognized object types within the system being integrated.

(3) Register Glossaries/Thesauri: DLII’s lexical analysis engine uses a unified glossary of terms from participating collections and services to find key phrases associated with the glossary entries over the entire document or screen content. These become supplemental content-based links (not shown in Figure 1).

Most other systems could be integrated in the same manner as digital library collections and services.

**MUCH MORE THAN LEXICAL ANALYSIS**

DLII generates the majority link anchors and links automatically through structural analysis using relationship rules. (Lexical analysis supplements these structural links.) If a system can operate on an element, DLII generates a link leading directly to this system’s service. For example, given a discussion thread about a course, any time that course’s identifier appears, DLII automatically detects this and adds an anchor over the course identifier.

Relationship rules define which relationships (links) should be available for which kinds of elements. For example, in Figure 1, the relationship rule underlying the second link would include the following parameters:

- the element type (in this case “document”)
- the link display label (“Summarize document...”)
- relationship metadata (semantic type, keywords, etc., useful for filtering)
- the destination collection or service (in this case the “Arizona Document Summarizer”)
- the exact command(s) to send to the destination system (“\http://keats.com.arizona.edu:8080/ebizport/serlet/ebizport/Summarizer.jsp?url=X\&length=3” where X is the document URL)

- any relevant conditions for including this relationship (including access restrictions)

Because they operate at the “class” or “kind of element” level, each relationship rule works for every element of that class. E.g., the rule above applies to any “document” element found within any screen or document displayed.

Each relationship rule represents a single relationship for a single element class. As elements can have many relationships, each element class can have several relationship rules. Each element instance triggers the same set of relationship rules, assuming conditions are satisfied for each. In Figure 1, two relationship rules triggered for the “document” element (or more rules triggered, but DLII’s collaborative filtering produced this customized list).

**BENEFITS TO INTEGRATION**

Digital library systems should find several benefits to integrating through DLII:

- DLII virtually enlarges the size of a collection and the "feature set" or services that a system provides through links to related information and services.
- DLII brings more users to a system because it has related information or relevant services.
- DLII streamlines individual systems by providing direct access through links among a single system’s information and functions, sparing the user from navigating through a possible series of menus.

**CONTRIBUTIONS**

This research’s primary contribution is providing a relatively straightforward, sustainable approach for integrating information systems. At the core lies an architecture for “lightweight systems integration through linking.”

Other contributions include:

- Developing filtering mechanisms for customizing large sets of links to particular users.
- Combining automatically generated structural links and links found through lexical analysis as a way of achieving integration.

**ACKNOWLEDGMENTS**

We gratefully acknowledge support by the NSF under grants IIS-0135531 and DUE-0226075, and the UPS Foundation. DLII is part of the National Science Digital Library project (http://www.nsdl.org).

*Figure 2: DLII Architecture. DLII is within the shaded area. The dashed paths indicate that once integrated, collections and services can share features through DLII links automatically. Integrated systems also continue to operate independently of DLII.*