Peter W. Martin
Legal Information Institute
Jan. 18, 1996

Some Reflections on Delivering the Full-Featured Lexis Services Via an HTML Gateway

I. A Brief Review of the Argument for Using HTML to Deliver Lexis-Nexis Services Followed by Some Disclaimers

The argument for delivering LEXIS-NEXIS services using future generation of Web browsers (and word-processors that operate as browsers) relates closely to the case for LEXIS-NEXIS not attempting to build a legal word-processor. Given the vast potential demand other parties are investing far more than LEXIS-NEXIS could ever hope to to develope access software. Long term it makes little sense to attempt to compete unless there is some distinctive quality of the data and data tools that comprise the LEXIS-NEXIS service that broader purpose access software will not deliver adequately. Furthermore, as more and more of the current and prospective LEXIS-NEXIS core customer base is drawn to the Internet for business, weather, travel information, as well as entertainment, news, and commercial activity, for LEXIS-NEXIS not to be accessible via that same pathway, using the same software will become a serious liability. Already Web browsers provide attractive user features that reflect unfavorably on the LEXIS-NEXIS session manager(s). The gap will inevitably widen. And the quality of the services will be tainted by user perceptions of the interface. The experience of working with first year law students as they gain their initial impressions of competing computer-based information systems leads me to be acutely conscious of how easy it is to conflate the access software with what it delivers.

Powerful additional arguments for an HTML gateway stem from all the other information sources and services assembling there. New customers and new competitors are coming on the Net at an enormous rate. LEXIS-NEXIS has assets that should serve it well in this rapidly changing venue -- unique content, comprehensiveness in important domains, rapid updates, high quality data (subject to quality review, fielded, formatted, hypertext linked), powerful search tools. In the “information bazaar” of the Internet such assets can be used to generate additional return. Kept apart from the Net they will increasingly be threatened by it.

II. Some Benchmark Gateway Applications

Many search engines have already been linked to the World Wide Web through HTML gateways. Two notable examples that, along with others, warrant attention as LEXIS-NEXIS considers how its services might be brought to and through the Web are the Folio Webserver and the “Thomas” site’s use of Inquery,

developed by the Center for Intelligent Information Retrieval at the University of Massachusetts, Amherst to provide search access to bills and acts of the current Congress. The former is particularly useful as a benchmark because it preserves both structure and hypertext in underlying data in addition to providing access, via the Net, to a sophisticated search engine.

III. Breaking the Exercise Into Pieces

A. Display, Files, Printing, and Interaction with the User

At one level LEXIS-NEXIS offers an easy case for HTML. Compared to the scholarly works of some other disciplines which make extensive use of graphic or tabular material on the one hand or format sensitive text and graphic reports, magazines on the other, the texts that LEXIS delivers are flat. In their native database format (VISF) more recent judicial opinions contain more format and structure than LEXIS shows (with such font and format features as italics and indent tagged) but they not rich to the point of creating a challenge to HTML. Screen by screen, the current LEXIS Session Manager user experience could be rendered in HTML. With most screens, the HTML could be generated on the fly. This should be true of those screens that are currently derived through a transaction with a data source. Some LEXIS screens, principally those that map "libraries" and "files" for user selection, are, I believe, stored as full screens. If so they would either require a more complicated parser or alternatively intact preservation via the HTML <PRE> tag. (The difference is between a screen with rows and columns that is generated from structured data and one that is held as a flat character string including a lot of spaces.) Assuming that all LEXIS screens are generated from data can be interpretted as HTML as readily as the character streams that govern display on the current session manager(s), it should be easier to translate LEXIS document and service displays to HTML than a high-end Folio VIEWS infobase.

The challenge of converting richly structured Folio Views infobases to HTML is discussed in the LII's working paper 94-4 (http://www.law.cornell.edu/papers/lii/fffhtml.htm). Any gateway generating HTML from Views on the fly will perforce leave value behind and is very likely to fail to make use of effective proxies that a more tailored porting process might deploy. (Compare the LII's Folio VIEWS version of Article 2 of the U.C.C. on the LEAP Library CD-ROM first with its rendering via the LII's Folio webserver [http://www2.law.cornell.edu/folio.pgi/UCC2?] and then with the LII's native HTML version [http://www.law.cornell.edu/ucc/2/overview.html].) By contrast HTML is capability of delivering substantially more value that the current LEXIS-NEXIS session manager(s). (This report takes the current session manager(s) as its reference for two reasons. First, referring to multiple future generations of the session manager(s) quickly becomes unwieldy. And, second, I know very little about plans for future versions of the software.

Translating session manager screen, by session manager screen, to HTML would, however, be unacceptable from both user and vendor perspectives. The typical web transaction involves transferring substantially more data than a session manager screen (text file sizes range well above 20K). Depending on the quality of the Net connection (on both client and server side) that transaction may well take longer than the painting of a fresh LEXIS-NEXIS screen, but once accomplished it frees the user of dependency on the server (and vice versa) so long as the user is browsing up and down that file or printing it (or any of those acquired previously in the same session or perhaps even a prior one).

The expectations (and related reponse time tolerance) generated by this mode of operating mean that a LEXIS-NEXIS HTML gateway would want to serve files significantly larger than a session manager screen. In the case of a wide variety of LEXIS-NEXIS user transactions this larger unit will correspond to the "document construct" currently employed. For example, in most cases, a set of search results displayed in CITE mode, a decision or article displayed in KWIC, a Shepard's or Autocite report on a case constitute logical units of information for the user that are not so large as to produce unacceptable response times. (Current browsers do, after all, allow the user to begin to scroll down a text document before it has been completely received.) Some "documents" will, however, exceed that threshhold. Some Supreme Court decisions with multiple opinions, major law journal articles, and many entries in the Federal Register would fall in this group. A reasonable approach to such instances would be to have a system limit on file transfer size. Documents exceeding the limit would be transferred in chunks much as the Folio webserver delivers portions of an infobase that carry links that will call for the next (and prior portions). The system might allow some degree of user configuration (based on band width or other considerations) subject to the system limit.

B. Maintaining Position in a Multi-Dimensional, Multi-Service Information Universe - Division of Labor Between Host and Client

There are no WWW sessions. The architecture is radically different from that of connection between the LEXIS-NEXIS user and host. In the latter case the host keeps track of what the user has done. Not of that progress and location information is stored by the client, at least automatically. As simple a move as a return to the immediately preceding screen or printing the current document depend on the host. With the Web information is passed in discontinuous transactions.

It would, no doubt, be possible technically to use a WWW gateway to construct "sessions" using a WWW gateway. A user would in such a system establish identity and begin a session which would continue through repeated discrete WWW transactions until the user (or some default mechanism) ended it by doing an "end session" transaction. The superficial attraction of such an approach is that all the sophisticated logging of user position and prior steps could be maintained through "sessions" -- available to users as a

ready set of options. The major drawback of this approach is that it throws away much of the value of WWW architecture. Web clients are designed to cache and display the results of previous steps (within limits as set by the user). Users can electively save important documents complete with all their hypertext functionality, holding them beyond the caching period. Alternatively users can bookmark documents so that a later return will draw in additions or other changes.

To deliver LEXIS-NEXIS through such a system is to alter it substantially. From a user's standpoint (leaving all issues of pricing aside, for the moment) the changes net out some loss of information about prior activity and current options against major gains in terms of control. And from a systems standpoint, letting go of keeping track of user sessions must represent major potential efficiency gains.

Consider two illustrative sessions.

1. A LEXIS-NEXIS user seeks the appellate and statutory authorities bearing on the liability of a dog owner for injuries caused by his animal in some particular state. Selection of "library" and "file", search entry and search results with options, working through a search set, following out-references from an individual document to other decisions or statutes and return, using citation services with regard to one or more opinons, with preservation of selected documents or passages in digital or print form (or both) can all be accomplished through a series of transactions in which the server passes both data and related action possibilities to the user. The total universe of potentially useful documents is relatively small. The number of steps from start to finish may be no more than a dozen, not counting the retrieval of the individual decisions or statutory sections the user is likely to want to retain for analysis and use.

2. A LEXIS-NEXIS user seeks the Federal decisions bearing on a spouse's vulnerability to the exercise of property forfeiture under 21 U.S.C. § 881(a)(7) against their drug dealing mate. The difference of this second scenario is that the number of potentially relevant decisions is far higher. As a consequence the trained LEXIS-NEXIS user is likely to engage in a highly iterative exploration process involving multiple searches of what amounts to the same or related answer sets.

C. Keeping Track of Use and Maintaining Security

The "session" concept has implications far beyond logging the user's tracks to permit his orientation and return. Historically, it has been at the heart of the metering and charging for LEXIS-NEXIS. Any shift to HTML delivery of full LEXIS-NEXIS services will require reinventing (and implementing) a new or revised coherent charging system. Even those elements of the current charging scheme that are tranaction based will have to be repackaged. Assuming that full case "documents" pass from server to client, retrieving (LEXSEE), browsing (reviewing online), and downloading or print are accomplished in that one transfer.