Overview

Seaglex Software has developed unique technologies for extracting, structuralizing and classifying data from the Internet and Intranets and for automating human interactions with Web sites. Our technologies include:

- **HTML Pattern Recognition**: allows users to convert arbitrary Internet/Intranet HTML sources into streams of structured XML data. It employs sophisticated pattern recognition algorithms to detect repetitive and recursive patterns with minimum help from the user. The parsing scripts generated at the end of the recognition process can then be automatically run to extract data from the HTML sources. These routines can extract data even if the HTML source appearance varies every time new data is retrieved. No manual programming is necessary; user interacts with the web application in very much the same fashion as browsing the web by pointing, clicking, and annotating.

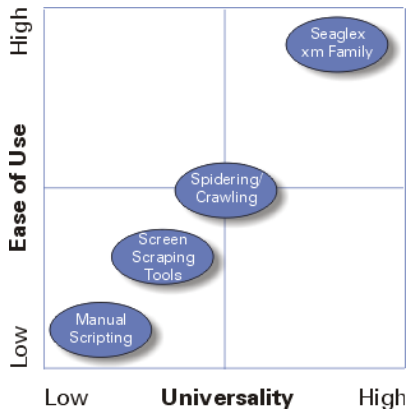
- **Playback/Recording Technology**: records user interactions with HTML sources so that user actions can then be replayed in repetitive cycles and on a massive scale. This technology also allows XML-driven automation of form input and secure login procedures. Recordings of user actions become parameterized procedures that can be called with different arguments through programmatic APIs. Again, user interacts with the application in the web-browsing fashion and is not required to develop any custom code.

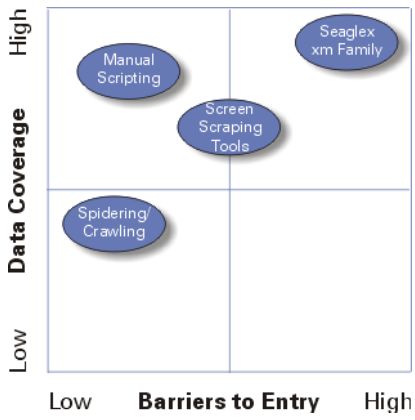
- **Creation of Uniform HTML Source Schemas**: by combining HTML pattern recognition with playback/recording technology, we produce uniform XML schemas that correspond to data contained in Internet/Intranet sites.

- **XML Schema Unification**: this crucial component of our technological platform allows unification and merging of XML data from disparate sources into one uniform stream of XML data. Our unification technology allows integration of data extracted from various HTML sources by Seaglex's technologies with external XML data.

Our key technological advantage is the universality and ease-of-use of our tools. According to Gartner Group, labor represents 80% of the time and cost to develop and implement today's web infrastructure. With Seaglex's products, this crucial investment of time and resources is drastically reduced.

Due to the variety of HTML techniques and frequent errors in "human"-generated HTML code, the task of creating a universal HTML-parsing tool, such as that designed by Seaglex, is highly complex. Seaglex Software managed to overcome the considerable barriers to entry and create truly universal, robust, and fully automated solutions. The following diagrams illustrate relative positioning of our technologies vis-à-vis traditionally used manual scripting, spidering/crawling, and automated screen scraping:





Seaglex xm-Series Development Tools

Based on the technologies outlined above, Seaglex Software has developed a set of tools for systems integration companies, technology-savvy vertical market organizations, and software vendors. These tools are also used by Seaglex development and professional services teams for creating and deploying custom vertical market applications. Seaglex xm-series product family is currently in the alpha testing stage with full product release scheduled for Q3 2002.

The key components of our product family are:

xmEditor – Design Time

- automation of Web site navigations
- automation of XML-driven form input and secure login procedures
- structuralization of HTML documents

xmUnifier – Design Time

- unification and merging of XML data

xmRunTime Engine and xmUnifier RunTime Engine

- unattended execution of scripts generated by xmEditor and xmUnifier

The following diagrams illustrate our design time and runtime architectures:

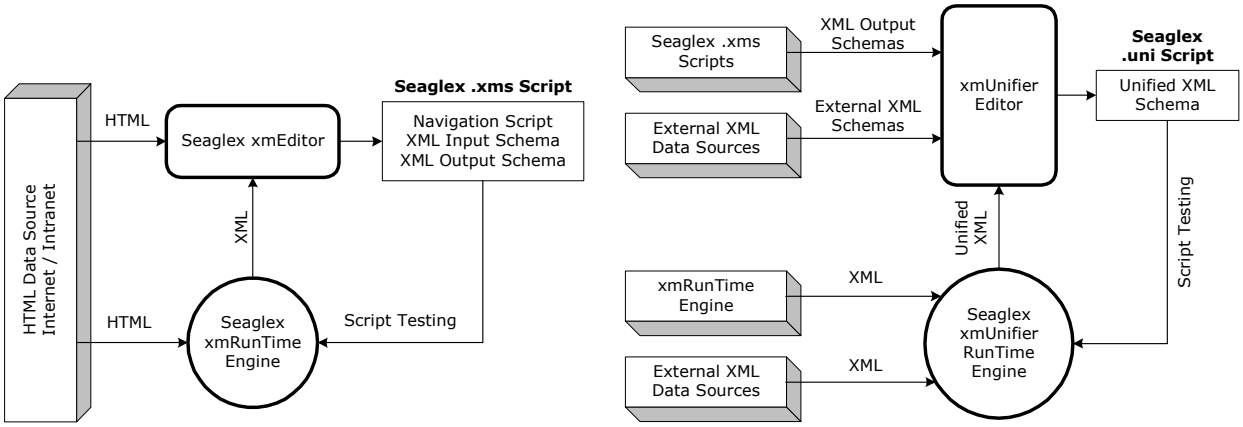


Figure 1. Design Time Architecture

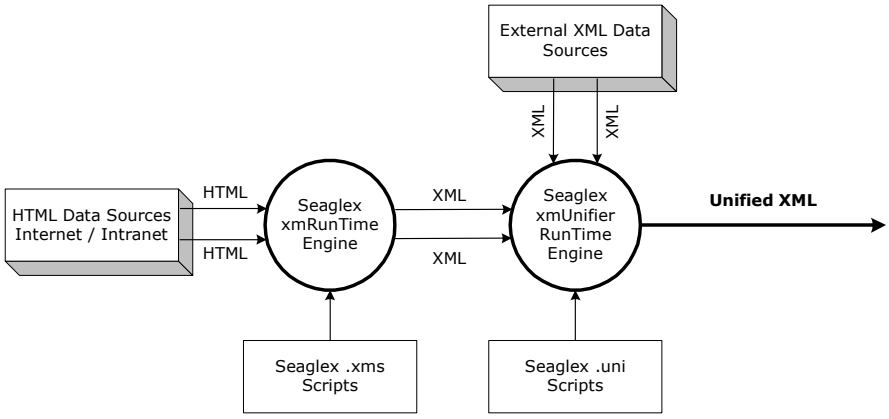


Figure 2. Runtime Architecture – Structured Data Extraction and Unification

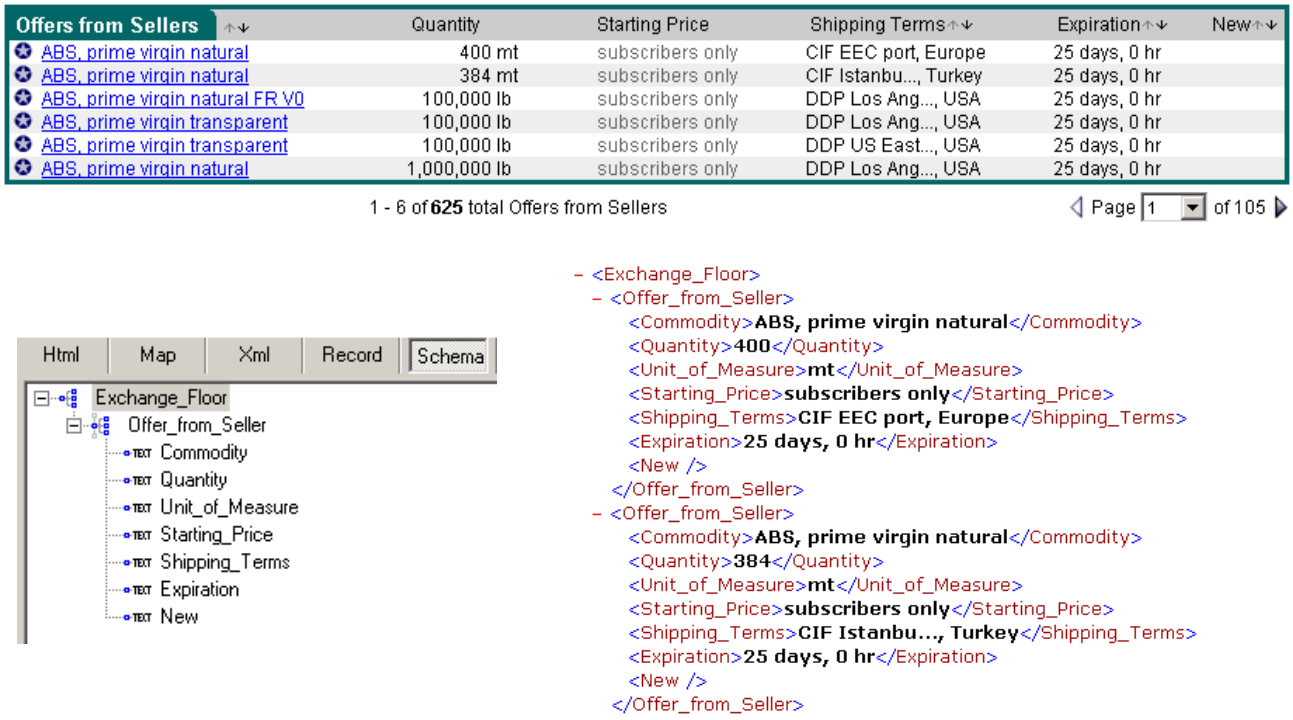
xmEditor

This is the key component of Seaglex offerings that comprises automation of web site navigations, XML-driven form input and secure login procedures, and structuralization of web pages resulting in XML data extraction. All of these features are described below. Once navigation, form and login information input, and web page structuralization processes are described via xmEditor's visual tools, resulting navigation scripts and input and output XML schemas are stored in Seaglex's proprietary .xms format. In order to test and fine-tune recorded procedures, xmEditor uses the xmRunTime Engine.

Mapping Web Sites to Extract Structured or Textual Information

xmEditor provides intuitive means for structuralizing HTML documents and presenting them as streams of XML data. This tool is universal and does not require any custom code development. xmEditor creates XML schemas that represent information contained in HTML documents; all extracted information is stored in XML format.

The following is an XML-structured representation of transactions listed on the World Chemical Exchange (www.chemconnect.com) as generated by xmEditor:



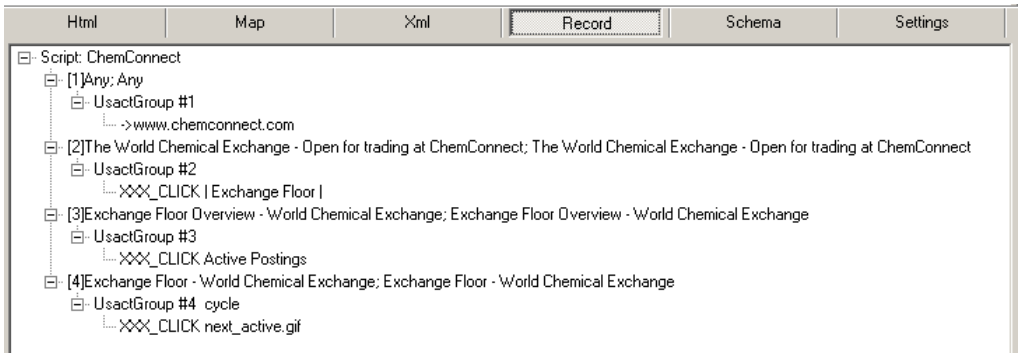
According to a survey conducted by Seaglex Software, structuralization of above web site with the use of currently available technologies requires a minimum of 15 billable developer hours. Moreover, manually created scripts require constant maintenance to reflect even the slightest changes in form structure. xmEditor allows non-skilled developers to create robust XML-driven scripts in minutes, which is a significant technological breakthrough that realizes considerable labor and cost savings.

Automating Interactions with Web Sites

In order to access data in a web site (or inside an Intranet repository), a user must navigate this site by following hyperlinks and viewing multiple pages of information. xmEditor allows fully visual and intuitive recording and playback of complex web site navigations. No custom programming is required.

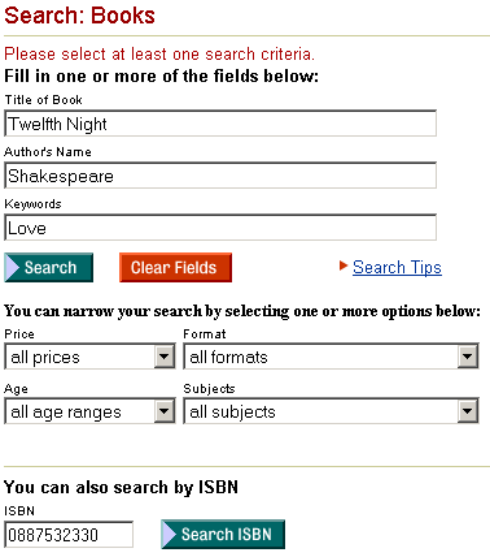
In the World Chemical Exchange example given above, in order to access the publicly available offers from commodity sellers, a user would enter "www.chemconnect.com," then follow the "Exchange Floor" link on the home page, then select "Active Postings," and cycle through 105 pages of listings to view all available information. These interactions would remain the same even if additional links or graphics were added to the pages preceding the data, or if the number of pages to cycle through increased or decreased. Since our scripts mimic human interactions with Web sites, we can provide robust and flexible solutions that do not require constant maintenance and troubleshooting.

The following script of user actions was automatically created to describe this navigation routine. No custom coding was involved; complete script creation took less than two minutes.



Automating Form Input and Secure Login Procedures

xmEditor structuralizes the form input process and presents all available fields as nodes of an XML schema, which enables automated XML-driven input procedures. The following example illustrates a search form schema generated by xmEditor that facilitates browsing Barnes&Noble.com:




```
- <DataIn>
    <Title_of_Book>Twelfth Night</Title_of_Book>
    <Author>Shakespeare</Author>
    <Keyword>Love</Keyword>
    <Price>all prices</Price>
    <Format>all formats</Format>
    <Age>all age ranges</Age>
    <Subject>all subjects</Subject>
    <ISBN>0887532330</ISBN>
  </DataIn>
```

xmUnifier

As most applications of our technology involve structuralization and normalization of data across multiple sites/documents, unification and merging of often disparate schemas of these various sources is crucial for processing extracted data. Moreover, for integration with external XML sources, such as backend business systems, it is essential to unify xmEditor-generated schemas with schemas provided by the external data sources.

xmUnifier provides powerful visual schema mapping tools and records resulting unification and merging procedures in its proprietary .uni format. In order to test and fine-tune recorded procedures, xmUnifier uses the xmUnifier RunTime Engine.

xmRunTime Engine and xmUnifier RunTime Engine

As illustrated in the runtime architecture diagram, xmRunTime Engine accepts .xms scripts as inputs and generates streams of XML-structured information from Internet or Intranet data sources that correspond to the scripts. xmRunTime Engine is also used for testing and fine-tuning .xms scripts in the xmEditor design environment.

xmUnifier RunTime Engine uses .uni scripts generated by xmUnifier and combines multiple streams of XML data (that are provided by xmRunTime Engine or external XML data sources) into a unified stream of information for further processing.

Backend Systems Integration

By utilizing Seaglex tools to structuralize and extract XML-formatted data from a variety of Internet and Intranet sites, our clients achieve seamless integration of data from these unstructured sources with existing backend business systems. An example of direct integration on data level appears below:

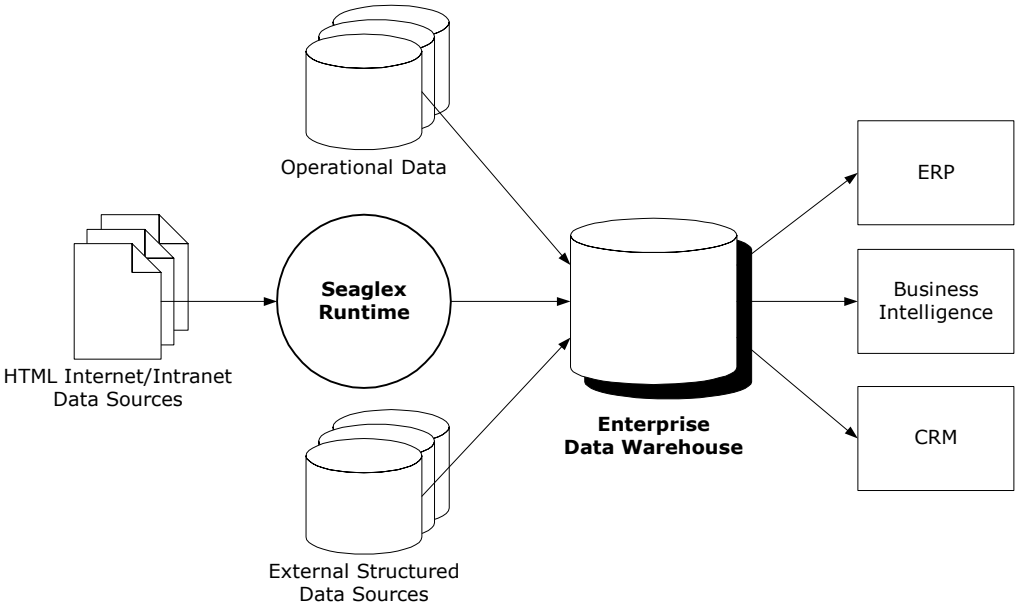


Figure 3. Example of Backend Integration

According to IDC, "if we can develop better ways to utilize unstructured content, we would have a powerful advantage for gaining more knowledge about our businesses and customers than our competitors can muster." Seaglex's products make this goal a reality.