TECHNOLOGY AUDIT

Lixto Online Market Intelligence

Lixto

BUTLER GROUP VIEW

ABSTRACT

Lixto’s Online Market Intelligence solution, which includes vertical offerings for the travel, consumer goods, and electronics industries, provides an automated way of extracting (primarily pricing data) from Web sites and repurposing it for notification, analytics, and dashboard reporting. Companies need to respond quickly to changes in competitor offerings but traditional market research methods cannot meet the speed and data coverage requirements of the online world. Lixto provides modules for visually defining data patterns within Web sites, extracting the data, and performing data analytics. Butler Group is impressed with the solution’s visual data-type definition process which allows non-technical users to quickly build up an exhaustive list of sources to be monitored. Also, the managed services option and Lixto’s domain focus lead to a reliable and low maintenance solution well aligned with the needs of the target sector.

KEY FINDINGS

- Highly visual data pattern definition.
- Automatic data collection and aggregation from large numbers of external Web sites.
- Requires manual configuration of connectors for each data source.
- Based on XML, SOAP, and J2EE.
- Provides industry-specific reporting, analytics and visualisation.
- Heuristics-based data extraction handles minor page changes and AJAX-based sites.
- Each data source requires manual navigation to identify data patterns.
- Utilises supervised learning approach.

LOOK AHEAD

Lixto is planning to add improved quality control mechanisms in order to validate gathered market data.
FUNCTIONALITY

Over the past decade the Internet has evolved to become a formidable vehicle of commerce, with a rich ecosystem of transactional Web sites, and sites that provide comparative information related to vendors and products, across a wide range of consumer categories. The ubiquity and real-time nature of the Internet has enabled and driven the need for quick, and in certain cases even real-time responses, to competitors’ actions in areas such as product and price revisions, and customisations. This need for agility is particularly true for items with a finite life, such as hotel rooms and air line tickets. For these sectors, while the Internet presents an embarrassment of riches in terms of freely available information, there is a lack of solutions that offer a reliable and automated way of aggregating the free information in a structured enough manner, to enable these real time product and price changes. A few industries, most notably the airline industry, have developed shared platforms for real-time price and product information exchange, to streamline transactions. However, for most industries across sectors and geographies, aggregating freely available information from the Web in an automated, exhaustive, and reliable way, coupled with analytics capability, is the way to go for short term price and product revisions, and for long term strategy formulation.

Product Analysis

Lixto, through the company’s Online Market Intelligence solution, addresses the aforementioned problem. The solution allows enterprise users to create connectors for specific web sites that extract required data (such as hotel room rates) from various online sources, convert the unstructured data into a structured format, and finally repurpose and present the data to the end-user in a suitable format for notification, analysis, and integration with in-house business applications. The data collected is cleansed and aggregated to generate dashboard views and detailed reports. The solution for the hotel industry, which enables analysis of hotel room rates, is the focus for this Technology Audit.

The solution can be used by travel industry players to gather competitor and market intelligence information from public Web sites; comparing prices to ensure their own pricing levels are appropriately set, for example. It can also be used to ensure pricing consistency within a distribution channel. In this scenario, a hotel could monitor its distribution channel Web sites to ensure that hotel room prices remain within a specified price range. Using the notification workflow capability, an alert can be configured in the event of a price change beyond a specified band. The detailed reporting capability can be used for analysing availability and price per competitor or per category. Lixto Online Market Intelligence is able to deliver business value by automating and bringing scale to an otherwise manual task.

Data extraction is based around data pattern recognition and Web page navigation. It involves drilling down into the Web page content structure, or more accurately the HTML tree structure, of the target Web site and specifying the class of data items, such as a list of prices to be extracted, in a highly visual way. The ‘class of data items’ is specified by establishing a set of rules (such as: select lines that begin with the word ‘Price’), and using the Visual Developer function that is included with the solution to select instances (such as highlighting five different pages areas that have prices of items), using mouse clicks rather than coding.
The extraction rule formulation process runs in parallel with the Web page navigation specification process. The navigation specification process involves specifying link traversals, form fill ups, and user log ons. The aforementioned processes run in parallel with the process of developing the data structure for the extracted data. The data structure could be imported from an existing structure or from a standard structure such as Really Simple Syndication (RSS). The navigation specification, ‘class of data item’ selection, and data structure building is performed using the Lixto Visual Developer product (based on Eclipse and the Mozilla browser engine). At the time of data extraction, the Lixto solution performs a heuristic analysis to retrieve all the data items that most closely match the specified rules. Runtime data extraction is performed by the Lixto Transformation Server. The Transformation Server extracts data based on configurable events, such as data retrieval schedules.

Retrieved data is cleansed and aggregated and can be repurposed in a variety of ways. The suite provides automated rate shopping and also offers end users a set of pre-defined, easily customisable reports and dashboards in order to provide an analytical overview of pricing trends. Reporting capabilities also include configurable alert notification and the option to interface with in-house business applications and business intelligence systems.

Butler Group is impressed with the highly visual way Lixto handles the data selection specification process, which ensures that the solution is suitable for use by marketing personnel (rather than IT specialists). Partly due to the heuristic approach, the solution can handle Web pages built with Rich Web Application technologies such as Asynchronous Javascript and XML (AJAX) and can tolerate minor changes in the page structure, such as appearance or positioning of page elements. In the event of erroneous data reporting, the class of data item specification can be tweaked using the Visual Developer tool. The solution is provided as a managed service, and can to a large extent handle the Web page changes on its own.

Overall, Butler Group is impressed with the solution and believes the Lixto approach would be relevant to companies in the hotel industry irrespective of size and regional focus. There are a few aspects that potential users should be aware of, however. Separate connectors have to be created and maintained for each data source which could become an issue if very large numbers of data sources are involved. Also, although the solution uses domain knowledge and heuristic intelligence to help with the process of defining data patterns and classes of data items, (and once established, data collection is automatic), Lixto Online Market Intelligence learns by example, so someone has to visit each site and navigate through the data in order to identify the data patterns and enable the creation of the data extraction rules (programs). Major changes to the site necessitate another iteration of the process. The positive side of this interactive learn-by-example, or supervised learning technique, is that it eliminates the time-consuming and expert task of creating and manually labelling a data model. It is also important to mention that Lixto currently focuses on the European market currently: this has implications for the limitations of the ‘out-of-the-box’ offering.

**Product Operation**

Lixto provides a solution that allows for rapid development of online market intelligence applications that provide a platform to analyse the product and price development in online markets. The architecture is as shown below in Figure 1:
The solution comprises the Lixto Online Market Intelligence Server, a data warehouse, and the Reporting and Analytics Cockpit. The Lixto Online Market Intelligence Server extracts product and price information from Web sites which are relevant for the specific industry segment, and it is this Web site data capture capability that differentiates the Lixto solution from other data capture and transformation applications. The server also performs data matching and cleansing operations on the extracted data. The data items are stored in a data warehouse for optimal reporting performance. The final component is the Market Intelligence Cockpit, which is an industry specific reporting and analytics application that enables analytical views of the market data. Unlike the Lixto Online Market Intelligence Server, this aspect of the technology is not unique but in Butler Group’s opinion combination plus the packaged nature and industry-specific capabilities make this a valuable feature. Alternatively, data can be used by in-house systems and reporting applications.

Lixto Online Market Intelligence is totally based on Java technology and standards like XML Schema, XSLT, SOAP, and J2EE. It uses the logic-based data extraction language ELOG in order to extract, transform, and deliver information from various semi-structured sources. Although this will be a new language to many developers and end users, there should be no need to interface directly with it as Lixto Online Market Intelligence is designed to be used by business users, not technical experts.

The solution has four modules, namely: i) the Web Harvesting module, ii) the Domain module, iii) the Analyse and Reporting module, and iv) the Workflow module as shown in Figure 2.
The Web Harvesting module is used for collecting information from various Web sites of the user’s choice using a Visual Extraction tool and an Execution Server. The Server extracts and translates all relevant information from HTML Web pages to a structured XML format that can be queried and processed by other programs. In the second step, generated XML data is processed by the runtime Lixto Transformation Server. Web data is retrieved automatically based on events, without the need for human interaction. The Transformation Server can also combine, transform, reformat, and deliver data. The whole process of modelling the workflow and dataflow is done in a graphical user interface in the Lixto Transformation Server. Graphical objects are used to symbolise components, and the flow of data and the workflow are graphically defined by drawing connecting arrows between these objects.

Within the Harvesting function, Lixto Visual Developer is used to visually define data extraction rules and patterns from sample web pages. As can be seen in Figure 3, the tool tracks Web page navigation. It also identifies properties and parameters to be used when gathering data. Users point and click to visually select relevant data items on the Web page – in this example the Check-in date field has been selected – which means data will be gathered from this field. Butler Group believes use of the graphical Visual Developer tool has the potential to deliver business value as it makes a complex technical process accessible to marketers and sales people.
The Domain Module is the second module in the Online Market Intelligence solution and its role is to perform the normalisation of the structured XML data and render it ready to be accessed by the end-user tool. Typically, the receiving system, upon receiving the XML data message, validates the data structure and stores the data package in a temporary data queue. The data is then extracted, transformed, and uploaded onto the data warehouse for end-user access, as is normal with extract, transform and load (ETL) programs. This module also provides a data model which fits the needs of the specific application domain that the user has initially defined. It also provides the facility for users to obtain details of matching similar products from various providers in the same market.

The Analyse and Reporting module provides an enterprise-class business intelligence platform. This module supports reporting, ad-hoc analysis, and alerting with the help of dashboards, reports, and scatter plots. The Management dashboards provide a summarised overview of the product landscape in addition to highlighting where the product is positioned vis-à-vis competition. The reports provide a detailed view on a given product’s pricing and availability, while the scatter plots provide users with analytical graphs highlighting data correlations such as price versus availability or price difference across competitors.

The Workflow Services module manages the automation of specific workflows, such as those triggered by market events. It also provides standardised interfaces to external software such as price and revenue management, which could be used for price and yield management, and rate compliance. Based on comparative pricing details of competitors obtained through the solution, organisations can make the necessary changes to their price points and other relevant parameters.
Apart from the data extraction from the HTML tree, the Lixto solution’s converter feature can be used for identifying data presentation patterns within unstructured formats such as from PDF, Excel, Word, or CSV, and the visual specification methodology can be used for data extraction.

**Product Emphasis**

In Butler Group’s opinion, Lixto focuses on five key aspects: 1) enabling data extraction specification by non-technical domain specialists, through a very impressive graphical adaptation of a declarative process that takes into account complex textual (lexical and semantic) and graphical elements of Web pages; 2) providing an automated and reliable way of extracting Web data through a heuristic approach; 3) providing an analytics platform that can incorporate specialised pricing capability; 4) providing the structured data to third party in-house applications and analytics tools; and 5) providing the solution as a managed service that drastically reduces in-house oversight. In Butler Group’s opinion, the combination of Lixto’s Web data harvesting capabilities, focus on key sectors such as the automotive and the travel/hotel industry, and the overall goal of being a low-touch out-of-the-box solution is fairly unique and presents an attractive proposition for companies in these sectors.

**DEPLOYMENT**

Lixto reports that non-technical users can use the Visual Developer tool but would need some knowledge of Web page structures. In addition, consulting help would be required for building the target data model. No major third-party system integrators are required for solution implementation unless the customer foresees an advanced integration with in-house systems. Lixto claims that the average time for implementation would be around three months and the solution can be deployed using a modular approach.

The solution is designed for ease of use by non-technical domain specialists (sales and marketing personnel) but data modelling (and related analysis) requires knowledge of the data and how to combine and manipulate it. This is something no solution can eliminate, so Butler Group believes organisations would need a data management/modelling specialist (either in-house or external consultant) during planning and deployment. Also, as one of the benefits is the ability to enrich existing internal information with external market intelligence, integration with internal systems will be an important means of generating ROI.

The Lixto Online Market Intelligence solution is a fully serviced solution provided by Lixto. It can be deployed on site and maintained by Lixto remotely or hosted in the Lixto data centre and be accessed via a Web interface. For integration with in-house systems, Lixto provides an XML-based Application Programming Interface (API) and a set of connectors.

The solution can be deployed with server operating systems such as Windows (2000, XP Professional, Vista), and Linux (Red Hat and SuSE). The corresponding client component runs on a Web browser which can be Internet Explorer or Mozilla Firefox. The solution uses third party solutions such as database software from Oracle Database, Oracle Business Intelligence, and Sun Glassfish Application Server.

Lixto’s end-user and developer training program can be delivered either on-site or via the Web. Lixto provides technical support through hotlines, a trouble ticketing system, and e-mail. Legacy integration extends the product’s management capability by means of API (Webservice, Java RMI), XML data interchange, and Relational Database Management System (RDBMS) connectivity.
Lixto as a company targets client companies across sectors. However, as would be expected from a solution of this nature, the core technology is customised for specific sector-oriented offerings, such as hotel and travel management. Lixto has centered its marketing and sales efforts on enterprises with over EURO 300 million in revenues.

Lixto sells its products through a direct sales force. Lixto’s solution provider partners include: Freudenberg IT, itelligence, MHP, Prolytic, and Sensix GMBH. The list of technology partners include: IBS AG, IQ’s, MIS, Oracle, and Software AG. Lixto also partners with companies providing pricing and revenue optimisation software and services and such as PROS, Vendavo and IDeaS.

The Lixto solution’s licensing structure is both annual and perpetual. The average initial project value for Lixto is in the lower 6-digit Euro range, based on the amount of data processed and the Web data connectors deployed, with license and service costs accounting for 70% and 30% respectively of the total cost. The annual maintenance and support is 22% of standard software components. The maintenance costs are substantially higher than average and can presumably be attributed partly to the considerable Web surveying that Lixto has to perform to maintain the ‘out-of-the-box’ nature of the solution. They also cover service releases, patches, and second level support. While the pricing of the tool is based on the amount of data processed and number of Web data connectors deployed, the standard and optional parts of the Lixto tool are as follows:

**Standard Parts:**

- Lixto Online Market Intelligence Server.
- Lixto Online Market Intelligence Cockpit.
- Lixto Online Market Intelligence Control Panel.
- Lixto Web Data Connectors.

**Optional Parts:**

- Lixto Reporting Server.
- Lixto Custom Workflow Support.

Lixto is planning to greatly enhance quality control mechanisms to enable validation of aggregated market data. In Butler Group’s opinion, Lixto addresses a critical and large niche in the field of market research, plagued by a plethora of free online information without a reliable and automated method of real time reporting and analysis. In the automated Web-based data aggregation market, Lixto has few direct competitors. However in Butler Group’s opinion the company faces competition from multiple quarters in varied forms. The offshore Business Process Outsourcing (BPO) industry has been very efficient at building scalable processes and leveraging low cost talent in an expanding list of locations. Amazon.com has launched the ‘Mechanical Turk’ service line, an online platform that enables companies to procure part time workers, who bid for projects. The service is low cost, and many companies have been sourcing list compilation services. While Lixto solution is certainly more elegant, the company’s continued relevance is contingent upon enhancing technology-based mechanisms and service partnerships that ensure quality, exhaustiveness, and depth of data provided. Nevertheless, Lixto’s pipeline of initiatives suggests that the company is on the right track towards enabling what is essentially a new way of delivering market research.
COMPANY PROFILE

Lixto Software GmbH is a spin off from the technical University in Vienna, EC3 (Electronic Commerce Competence Centre). Founded in 2001 the privately held company is headquartered in Vienna, Austria and also has offices in Germany. A team of computer scientists, mathematicians, and economists has developed the technology which is based on the idea of structured data delivery and syndication that enables users to extract relevant data from the Internet, delivered in a structured and customised way, without time consuming manual research.

Lixto has 35 employees with 90% based in Austria, and the remaining 10% in Germany. The employee base is segmented by function as follows: Research and Development – 35%; Sales and Marketing – 25%, Support and Services – 30%; and Administration – 10%. Lixto has 40 distinct customers across its entire product range, of which 10 have implemented the Lixto Online Market Intelligence solution. The referenceable client for the travel industry solution is hotel.de (travelers.com in UK market).

SUMMARY

Lixto’s solution is geared towards enabling quick responses to variations in competitor offerings, and towards serving as the single reliable repository for pricing and related information. Lixto’s data pattern definition process allows client organisations to quickly build up an exhaustive database of target sources, and also to quickly address changes in the target Web pages. The company’s domain focus and the outsourced services option helps in ensuring that data is exhaustive and reliable. Lixto presents a fairly unique proposition and is best suited for companies whose core business it is to make updated and accurate information available, such as portal providers. The solution is also extremely relevant to other business entities that need to react quickly to competition, which is often spread across thousands or even ten of thousands of separate commercial entities, a breadth that tests the limits of traditional market research.

<table>
<thead>
<tr>
<th>Table 1: Contact Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lixto Software GmbH</strong></td>
</tr>
<tr>
<td>A-1040 Vienna, Austria</td>
</tr>
<tr>
<td>Favoritenstrasse 16</td>
</tr>
<tr>
<td>Tel: +4312051224</td>
</tr>
<tr>
<td>Fax: +4312051224 12</td>
</tr>
</tbody>
</table>

For more information on Butler Group’s Subscription Services please contact one of the local offices above.