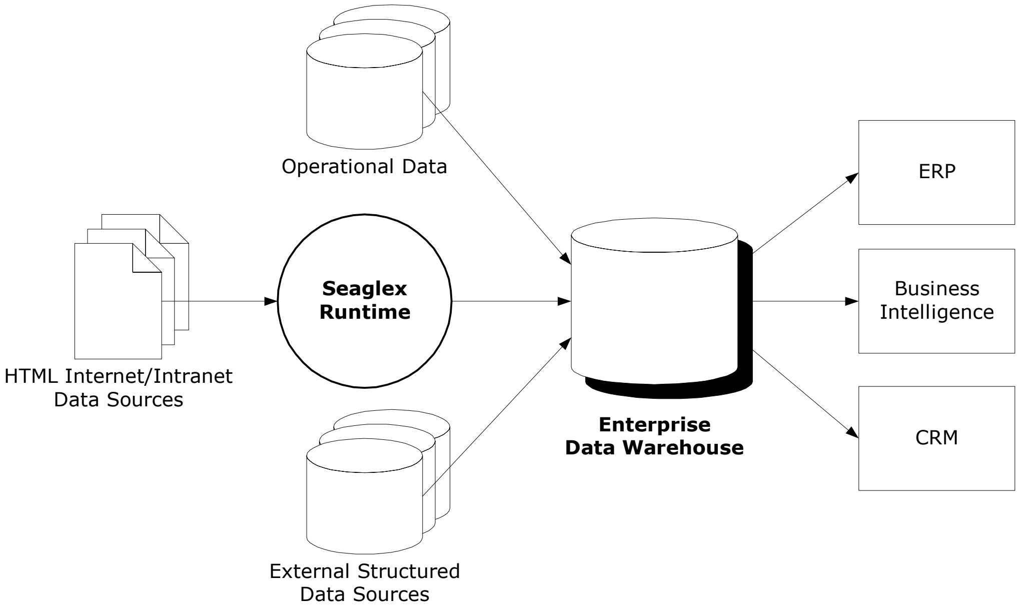
Overview

Seaglex Software is a leading developer of products for systematic location, identification, classification, and extraction of structured and unstructured (textual) information from Internet sources for business intelligence and enterprise application integration purposes. Our solutions are intended for industry analysts, market researchers, business and competitive intelligence consultants, and other knowledge workers that use information from the Web in their daily tasks.

Seaglex capitalizes on the deficiencies of current information management offerings and delivers turn-key solutions that drastically reduce time and effort spent by corporations on locating, integrating, and processing free “open Web” and paid-for content. 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 solutions make this vision a reality.

Our professional services team develops and deploys custom solutions that integrate data from the Web with your company database or business system. As a result, you can query and process Web-based data just like the information generated within your company.

This functionality is made possible by our revolutionary technologies that convert arbitrary web pages, HTML documents, or entire Internet sites into sources of eXtensible Markup Language (XML)-structured data. XML is an accepted industry standard that allows easy integration of extracted information with existing corporate data stores (Microsoft SQL Server, Microsoft Access, Microsoft Exchange, Oracle, Informix, Lotus Notes, etc.) and backend business systems, such as those offered by Baan, J.D. Edwards, Microsoft, Oracle, PeopleSoft, SAP, and others. The following diagram illustrates this concept:


Operational Data
HTML Internet/Intranet Data Sources
Seaglex Runtime
External Structured Data Sources
Enterprise Data Warehouse
ERP
Business Intelligence
CRM

The following sections of this white paper provide examples of our technology applications. Whereas not all of the illustrated techniques may be applicable to a particular client’s need, the breadth of our technology coverage ensures that we can address even the most challenging tasks, and the automated and visual features of our tools allow our professional services team to quickly deliver very advanced applications that normally are significantly more expensive and take much longer to develop and deploy.

Installation Requirements

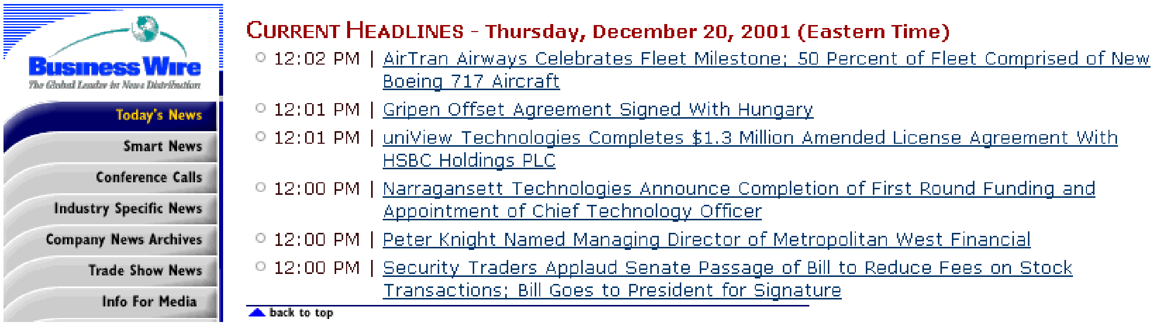
Our Web harvesting solutions have the following installation requirements:

- Intel Pentium III–compatible processor minimum recommended.
- 256 megabytes (MB) of RAM minimum recommended.
- Microsoft Windows® 2000 Server or Windows 2000 Professional operating system.
- Microsoft Internet Explorer 6.
- High-speed Internet access (ISDN or better) recommended.

Web Harvesting Technologies

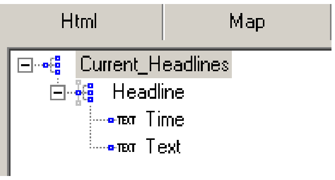
Mapping Web Sites to Extract Structured or Textual Information

Our HTML pattern recognition technology enables structuralizing HTML documents and presenting them as streams of XML data without any custom code development. The following is an XML-structured representation of headline news from Business Wire as generated by our tools:



CURRENT HEADLINES - Thursday, December 20, 2001 (Eastern Time)

○ 12:02 PM | AirTran Airways Celebrates Fleet Milestone; 50 Percent of Fleet Comprised of New Boeing 717 Aircraft
○ 12:01 PM | Gripen Offset Agreement Signed With Hungary
○ 12:01 PM | uniView Technologies Completes $1.3 Million Amended License Agreement With HSBC Holdings PLC
○ 12:00 PM | Narragansett Technologies Announce Completion of First Round Funding and Appointment of Chief Technology Officer
○ 12:00 PM | Peter Knight Named Managing Director of Metropolitan West Financial
○ 12:00 PM | Security Traders Applaud Senate Passage of Bill to Reduce Fees on Stock Transactions; Bill Goes to President for Signature



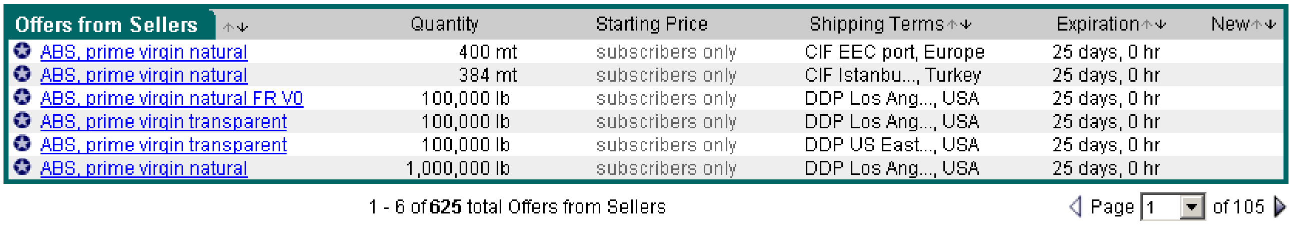
```
- <Current_Headlines>
  - <Headline>
      <Time>12:02 PM</Time>
      <Text>AirTran Airways Celebrates Fleet Milestone; 50 Percent of Fleet Comprised of New
        Boeing 717 Aircraft</Text>
    </Headline>
  - <Headline>
      <Time>12:01 PM</Time>
      <Text>Gripen Offset Agreement Signed With Hungary</Text>
    </Headline>
  - <Headline>
      <Time>12:01 PM</Time>
      <Text>uniView Technologies Completes $1.3 Million Amended License Agreement With HSBC
        Holdings PLC</Text>
    </Headline>
  - <Headline>
      <Time>12:00 PM</Time>
      <Text>Narragansett Technologies Announce Completion of First Round Funding and
        Appointment of Chief Technology Officer</Text>
    </Headline>
  - <Headline>
      <Time>12:00 PM</Time>
      <Text>Peter Knight Named Managing Director of Metropolitan West Financial</Text>
    </Headline>
  - <Headline>
      <Time>12:00 PM</Time>
      <Text>Security Traders Applaud Senate Passage of Bill to Reduce Fees on Stock Transactions;
        Bill Goes to President for Signature</Text>
    </Headline>
  </Current_Headlines>
```

When dealing with collections of links as in the above example, we can extract underlying HTML information to make the links “clickable” when they are delivered to a client and preserve the original “look and feel”:

```
- <Headline>
  + <![CDATA[   ]]>
    <Time>12:02 PM</Time>
  - <![CDATA[
        <span xmContextUrl='http://www.bizwire.com/cgi-bin/dh.cgi' xmContextTitle='Today_s News on the
        Net from Business Wire'>
        <TD><FONT face=Verdana, size=-1 sans-serif Helvetica, Arial,>
        <A href="http://www.businesswire.com/cgi-bin/f_headline.cgi?day0/213540203&amp;ticker=aai">
        AirTran Airways Celebrates Fleet Milestone; 50 Percent of Fleet Comprised of New Boeing 717 Aircraft
        </A></FONT></TD>
        </span>
    ]]>
    <Text>AirTran Airways Celebrates Fleet Milestone; 50 Percent of Fleet Comprised of New Boeing 717
      Aircraft</Text>
  </Headline>
```

12:02 PM	AirTran Airways Celebrates Fleet Milestone; 50 Percent of Fleet Comprised of New Boeing 717 Aircraft
12:01 PM	Gripen Offset Agreement Signed With Hungary
12:01 PM	uniView Technologies Completes $1.3 Million Amended License Agreement With HSBC Holdings PLC

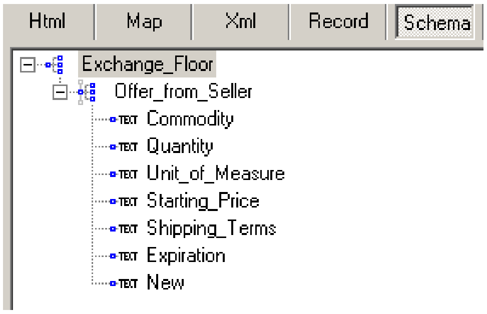
The following is a more complex example of structured data extraction from the World Chemical Exchange (www.chemconnect.com):



Offers from Sellers	Quantity	Starting Price	Shipping Terms	Expiration	New
ABS, prime virgin natural	400 mt	subscribers only	CIF EEC port, Europe	25 days, 0 hr	
ABS, prime virgin natural	384 mt	subscribers only	CIF Istanbu..., Turkey	25 days, 0 hr	
ABS, prime virgin natural FR V0	100,000 lb	subscribers only	DDP Los Ang..., USA	25 days, 0 hr	
ABS, prime virgin transparent	100,000 lb	subscribers only	DDP Los Ang..., USA	25 days, 0 hr	
ABS, prime virgin transparent	100,000 lb	subscribers only	DDP US East..., USA	25 days, 0 hr	
ABS, prime virgin natural	1,000,000 lb	subscribers only	DDP Los Ang..., USA	25 days, 0 hr	

1 - 6 of 625 total Offers from Sellers

Page 1 of 105



```
- <Exchange_Floor>
  - <Offer_from_Seller>
      <Commodity>ABS, prime virgin natural</Commodity>
      <Quantity>400</Quantity>
      <Unit_of_Measure>mt</Unit_of_Measure>
      <Starting_Price>subscribers only</Starting_Price>
      <Shipping_Terms>CIF EEC port, Europe</Shipping_Terms>
      <Expiration>25 days, 0 hr</Expiration>
      <New />
    </Offer_from_Seller>
  - <Offer_from_Seller>
      <Commodity>ABS, prime virgin natural</Commodity>
      <Quantity>384</Quantity>
      <Unit_of_Measure>mt</Unit_of_Measure>
      <Starting_Price>subscribers only</Starting_Price>
      <Shipping_Terms>CIF Istanbu..., Turkey</Shipping_Terms>
      <Expiration>25 days, 0 hr</Expiration>
      <New />
    </Offer_from_Seller>
```

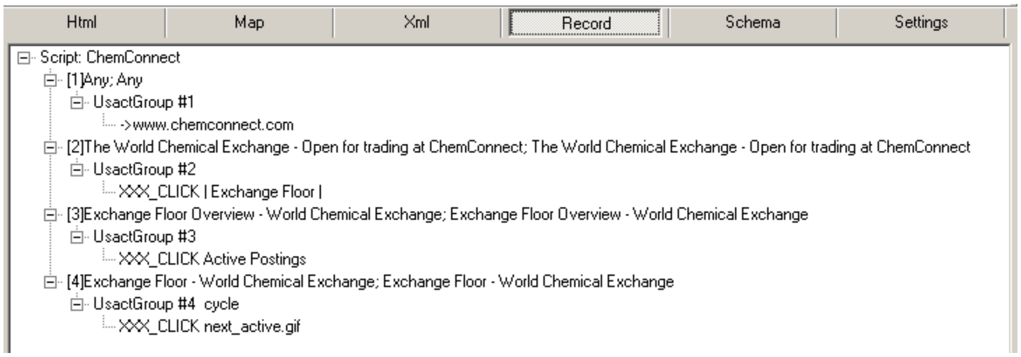
According to a survey conducted by Seaglex Software, structuralization of above web site with the use of currently available technologies requires a minimum of 15 developer hours. This traditional approach requires highly skilled engineers that typically charge upwards of $150/hour. Moreover, manually created scripts require constant maintenance to reflect even the slightest changes in HTML structure. Our technologies allow very rapid development of robust XML-driven scripts, thus enabling us to realize 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. Our technologies allow fully visual and intuitive recording and playback of complex Web site navigations.

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 press the “Next” arrow button to cycle through 105 pages of listings and 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.

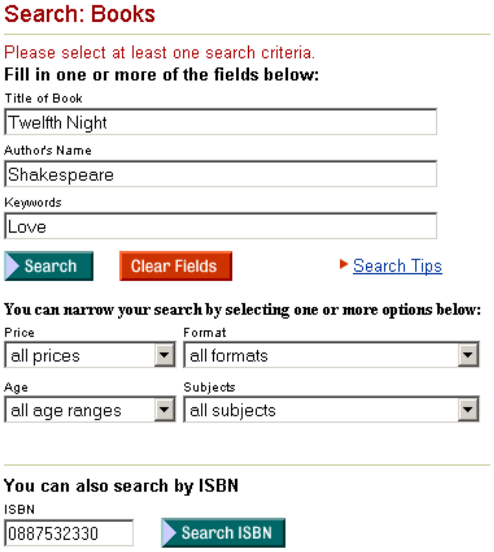
The following script of user actions was automatically created to describe this navigation routine:



```
⊟ Script: ChemConnect
   ⊟ [1]Any; Any
      ⊟ UsactGroup #1
         ->www.chemconnect.com
   ⊟ [2]The World Chemical Exchange - Open for trading at ChemConnect; The World Chemical Exchange - Open for trading at ChemConnect
      ⊟ UsactGroup #2
         XXX_CLICK | Exchange Floor |
   ⊟ [3]Exchange Floor Overview - World Chemical Exchange; Exchange Floor Overview - World Chemical Exchange
      ⊟ UsactGroup #3
         XXX_CLICK Active Postings
   ⊟ [4]Exchange Floor - World Chemical Exchange; Exchange Floor - World Chemical Exchange
      ⊟ UsactGroup #4  cycle
         XXX_CLICK next_active.gif
```

Automating Form Input and Secure Login Procedures

Our technologies allow structuralizing the form input process and presenting all available fields as nodes of an XML schema, which enables automated XML-driven input procedures. The following example illustrates automating search form input for Barnes&Noble.com:



Search: Books

Please select at least one search criteria.

Fill in one or more of the fields below:

Title of Book: Twelfth Night

Author's Name: Shakespeare

Keywords: Love

You can narrow your search by selecting one or more options below:

Price: all prices

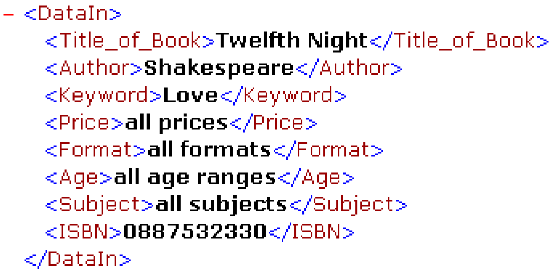
Format: all formats

Age: all age ranges

Subjects: all subjects

You can also search by ISBN

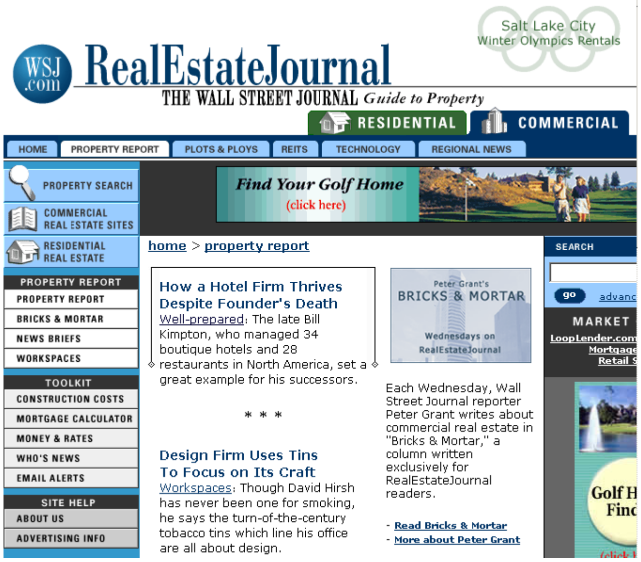
ISBN: 0887532330



```
- <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>
```

Filtering and Classifying Textual Information

When dealing with unstructured sources of information, we combine data extraction techniques with sophisticated filtration methods in order to detect and extract contiguous "article-like" textual content and distinguish it from collections of links, common HTML page elements, and other irrelevant "noise." In the following examples, only the colored article on the right will be extracted, while the page on the left and non-article elements on the right, including inline graphics, will be rejected:



How a Hotel Firm Thrives Despite Founder's Death

Well-prepared: The late Bill Kimpton, who managed 34 boutique hotels and 28 restaurants in North America, set a great example for his successors.

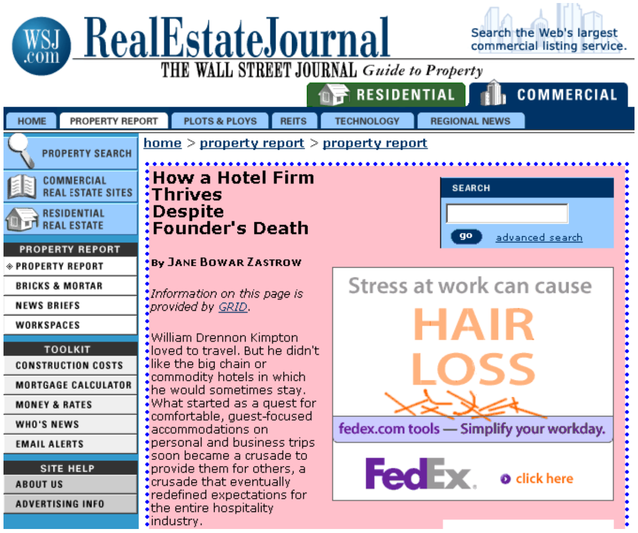
* * *

Design Firm Uses Tins To Focus on Its Craft

Workspaces: Though David Hirsh has never been one for smoking, he says the turn-of-the-century tobacco tins which line his office are all about design.

Each Wednesday, Wall Street Journal reporter Peter Grant writes about commercial real estate in "Bricks & Mortar," a column written exclusively for RealEstateJournal readers.

- Read Bricks & Mortar
- More about Peter Grant



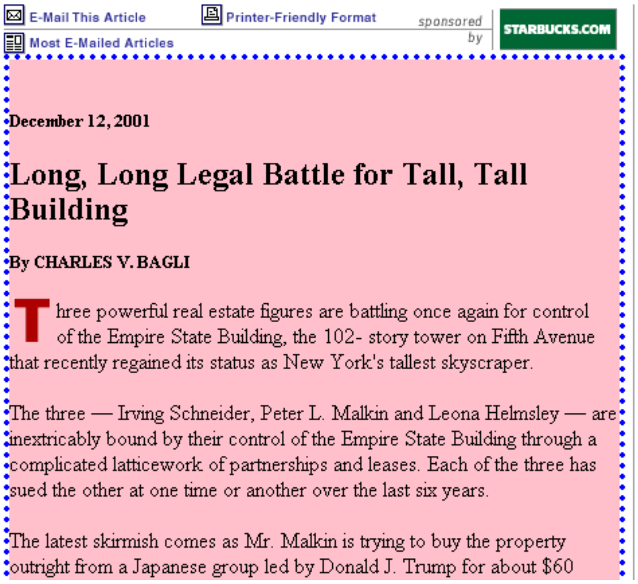
How a Hotel Firm Thrives Despite Founder's Death

By Jane Bowar Zastrow

Information on this page is provided by GRID.

William Drennon Kimpton loved to travel. But he didn't like the big chain or commodity hotels in which he would sometimes stay. What started as a quest for comfortable, guest-focused accommodations on personal and business trips soon became a crusade to provide them for others, a crusade that eventually redefined expectations for the entire hospitality industry.

Since even the information located in the "right" section of a newspaper or company site may be totally irrelevant, further content filtration should be applied to ensure very narrow focus of delivered data. As an example, The New York Times' real estate section contains mostly residential and occasional commercial real estate articles; for a commercial real estate analyst, all residential real estate news articles are "noise" that should be filtered out. We use Naïve Bayesian statistical algorithms and other machine learning techniques to distinguish between relevant textual content and "noise." In the snapshots below, our commercial real estate filter accepted the article on the left and rejected the article on the right:



December 12, 2001

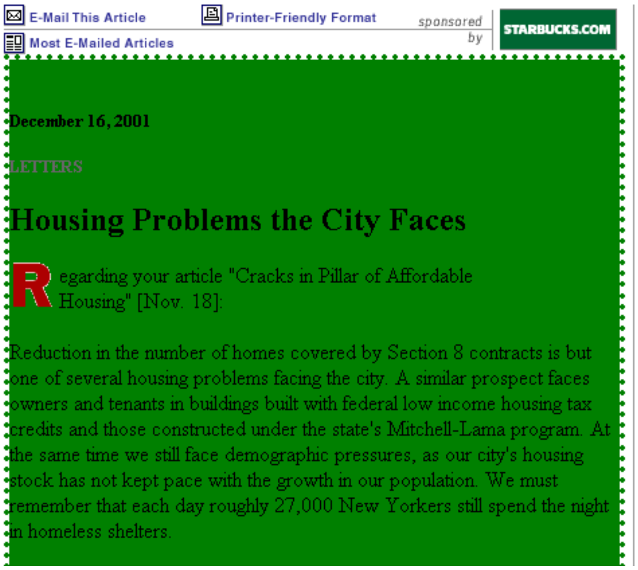
Long, Long Legal Battle for Tall, Tall Building

By CHARLES V. BAGLI

Three powerful real estate figures are battling once again for control of the Empire State Building, the 102- story tower on Fifth Avenue that recently regained its status as New York's tallest skyscraper.

The three — Irving Schneider, Peter L. Malkin and Leona Helmsley — are inextricably bound by their control of the Empire State Building through a complicated latticework of partnerships and leases. Each of the three has sued the other at one time or another over the last six years.

The latest skirmish comes as Mr. Malkin is trying to buy the property outright from a Japanese group led by Donald J. Trump for about $60



December 16, 2001

LETTERS

Housing Problems the City Faces

Regarding your article "Cracks in Pillar of Affordable Housing" [Nov. 18]:

Reduction in the number of homes covered by Section 8 contracts is but one of several housing problems facing the city. A similar prospect faces owners and tenants in buildings built with federal low income housing tax credits and those constructed under the state's Mitchell-Lama program. At the same time we still face demographic pressures, as our city's housing stock has not kept pace with the growth in our population. We must remember that each day roughly 27,000 New Yorkers still spend the night in homeless shelters.