The Amorphic system has been used to extract data relevant to the study of online auctions. IN THE FUTURE, SIMILAR SYSTEMS CAN BE USED TO EXTRACT DATA related to financial markets, online travel, or benefits, just to name a few.

domain. Since the domain ontology allows more than one set of keywords and more than one pattern to be specified, it can be used to extract data from several different Web sites in the auction domain. Once appropriate domain ontologies were created, the Amorphic information extraction system was used to extract data from 1,609 search-results pages and 626 single-item pages from the eBay, Yahoo, and Amazon online auction sites. Table 1 shows the prototype Amorphic system showed excellent performance for three Web sites tested.

To test the wrapper recovery procedures the Amorphic system was used to extract data from six additional online auction sites: Bidz.com; uBid.com; DellAuctions.com; CompUSAAuctions.com; BidVille.com; and ZBestOffer.com. The prototype Amorphic system demonstrated the ability to adapt to six additional Web sites it was not originally designed to support. The testing of the six additional auction sites did not require changes to the Amorphic program or the online auction domain ontology. Table 2 shows the information extraction results both with and without wrapper recovery. It shows the Amorphic agent was able to extract substantially more information from the six new Web sites using automatic wrapper recovery.

Conclusion

The use of external information for business decision making is not new. What is new is the abundance of information freely available via the Internet. However, this information is not being systematically included in current decision-making applications [8]. This research demonstrates that it is possible to reliably extract Web information for use in Web business intelligence applications. It will be possible for organizations to use a system like Amorphic to extract information of interest from Web pages for a wide variety of domains. These potential business intelligence applications will allow a deep and detailed look at small portions of the Web relevant to specific domains.

The Amorphic system has been used to extract data relevant to the study of online auctions. In the future, similar systems can be used to extract data related to financial markets, online travel, or benefits, just to name a few. Use of an information extraction system, like Amorphic, has the potential to provide businesses with access to up-to-date, comprehensive, and ever-expanding information sources that can in turn help them make better strategic decisions. C

References

1. Arasu A. and Garcia-Molina, H. Extracting structured data from Web pages. A*CM SIGMOD Record* (June 2003), 337–348.
2. Chidlovskii, B. Automatic repairing of Web wrappers by combining redundant views. In *Proceedings of IEEE Conf. Tools with AI* (Nov. 2002), 399–406.
3. Cohen, W., Hurst, M., and Jensen, L. A flexible learning system for wrapping tables and lists in HTML documents. In *Proceedings of the Conf. on WWW* (2002), 232–241.
4. Embley, D., Campbell, D., Smith, R., and Liddle, S. Ontology-based extraction and structuring of information from data-rich unstructured documents. In *Proceedings of the Conf. on Info. and Knowledge Management* (Nov. 1998), 52–59.
5. Embley, D.W., Jiang, Y., and Ng, Y.K. Record-boundary discovery in Web documents. *ACM SIGMOD Record 28*, 2 (June 1999), 467–478.
6. Gregg, D. and Walczak, S. Exploiting the Information Web. *IEEE Trans. on System, Man and Cybernetics Part C* (forthcoming 2006).
7. Knoblock, C., Leramn, K., Minton, S., and Muslea, I. Accurately and reliably extracting data from the Web: A machine learning approach. *Bulletin IEEE Computer Society Technical Committee on Data Engineering* 23, 4 (2000), 33–41.
8. Kushmerick, N., Weld, D., and Doorenbos, R. Wrapper induction for information extraction. In *Proceedings of the Conf. on AI* (1997), 729–735.
9. Laender, A.H.F., Ribeiro-Neto, B.A., da Silva, A.S., and Teixeira, J.S. Surveys: A brief survey of web data extraction tools. *ACM SIGMOD Record 31*, 2 (June 2002), 84–93.
10. Lerman, K., Minton, S., and Knoblock, C. Wrapper maintenance: A machine learning approach. *J. of AI Research 18* (Feb. 2003), 149–181.
11. Muslea, I., Minton, S., and Knoblock, C. A hierarchical approach to wrapper induction. In *Proceedings on Autonomous Agents* (1999), 190–197.
12. Srivastava J. and Cooley, R. Web business intelligence: Mining the Web for actionable knowledge. *J. on Computing 15*, 2 (2003), 191–207.

Dawn G. Gregg (dawn.gregg@cudenver.edu) is an assistant professor of information systems management in the Business School at the University of Colorado at Denver and Health Sciences Center.
Steven Walczak (swalczak@carbon.cudenver.edu) is an associate professor of information systems management in the Business School at the University of Colorado at Denver and Health Sciences Center.