Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

Universal Parsing Agent™ (UPA)

It seems that data often spends more time being reformatted than being analyzed to support decision-making. Universal Parsing Agent™ (UPA) was developed by Pacific Northwest National Laboratory (PNNL) as an innovative and unique approach to the problem of information extraction, transformation, and delivery. UPA accepts multiple datasets or streams of information, extracts information needed by users, and delivers results in their most useful form. UPA provides a flexible, reliable, and scalable solution to information workers' needs in today's high-volume dynamic data environments.

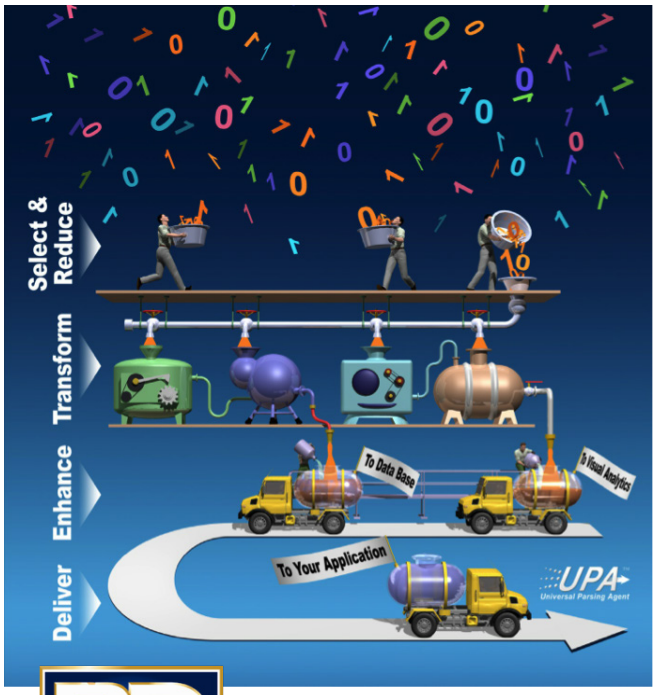
CHALLENGE

Information workers in government, business and academia struggle daily with formatting and structuring data for applications. Using available data extraction tools, they often spend more time cleaning, sorting, and reformatting data in preparation for analysis than analyzing the data itself. Data often arrives in massive amounts from different sources in various formats, further compounding the problem.

In addition, data extraction tools are often created for a single purpose. When data formats change, users cannot make modifications to the extraction tool easily. Instead, they must rely entirely on programmers to make the necessary changes. Even then, attempts to revise the software can become complicated and cost-prohibitive.

SOLUTION

UPA provides a reusable tool to identify and extract important elements from semi-structured and unstructured text documents, enhance those extractions with semantically descriptive tags, and generate output in the required format and structure. UPA empowers users to specify the data they need, instead of relying on programmers. Users tell UPA the exact information they want from each data source by creating a flexible



R&D Magazine named UPA one of the world's 100 most significant technical innovations in 2007.

template that describes this information. UPA templates are the software's directions for what data to find, extract, change, enhance, and output.

UPA's model allows the templates to act as instructions that a researcher would give to a research assistant; they describe what is needed, how it should be presented, "must-haves," alternatives, and even "would-likes." UPA is the research assistant that knows how to sift through the information to find what is needed on an immense scale. Further, when data sources such as electronic reports, web pages, and streaming data change formats, simple template modifications handle the updates easily.

IMPACT

UPA's capabilities are evident in many areas. With its processing speed of over 100,000 documents per hour, UPA can create unified data formats for research, prepare historical data for the Internet, dissect e-mail survey responses, and parse third-party online calendars. Organizations that must keep abreast of the competition use UPA to support their competitive intelligence needs by creating knowledge bases that reduce uncertainty and risk in decision-making.

ABOUT PNNL

Interdisciplinary teams at Pacific Northwest National Laboratory address many of America's most pressing issues in energy, the environment and national security through advances in basic and applied science. PNNL employs 4,300 staff, has an annual budget of nearly $1 billion, and has been managed for the U.S. Department of Energy by Ohio-based Battelle since the laboratory's inception in 1965.

For more information, contact:

Alex Gibson
Project Manager/Chief Architect
alex.gibson@pnnl.gov
(509) 372-6086

Dave Thurman
Manager, Analytics Programs
National Security Directorate
dave.thurman@pnnl.gov
(206) 528-3221

Pacific Northwest NATIONAL LABORATORY
U.S. DEPARTMENT OF ENERGY
Proudly Operated by Battelle *Since 1965*