

NEWSLETTER *InfraSoft* Die PROFIS

InfraSoft entwickelt Text-Konverter für APA

Mit Beginn des Jahres 2001 hat die *Austria Presse Agentur (APA)* ein Projekt begonnen, in dessen Rahmen die aktuellen Inhalte von rund 100 Zeitungen gesammelt und zentral zugänglich gemacht werden sollten. Zu diesem Zweck war vorgesehen, sowohl die redaktionellen Inhalte als auch die Anzeigen aller Zeitungen laufend in eine Datenbank zu übertragen und im Rahmen eines speziellen Service Kunden zur Verfügung zu stellen. Dazu ist es natürlich notwendig, die in sehr unterschiedlichen Formaten bei der APA eingehenden Daten vor dem Eintrag in die Datenbank zu vereinheitlichen. Ausgestattet mit dieser Anforderung hat der Generalunternehmer *Hewlett-Packard* die *InfraSoft* beauftragt, das Design eines entsprechenden Konverters für die APA vorzunehmen und im Anschluss die Software zu realisieren.

Offene Architektur

Da eine große Anzahl von Input-Quellen (ca.100 verschiedene Zeitungen) vorliegt, war der Wunsch nach einer flexiblen Lösung vordergründig. Der Konverter sollte zu einem späteren Zeitpunkt, wenn eventuell neue Datenlieferanten mit ggf. neuen Formaten dazu kommen würden, leicht zu erweitern sein. Auch auf der Output-Seite sollte der Konverter so flexibel wie möglich arbeiten. Könnte es doch sein, dass die zu füllende Datenbank einmal durch ein anderes Produkt abgelöst werden würde, das ein anderes Input-Format verlangt. Oder dass die durch den Konverter vereinheitlichten Daten

zum Beispiel in XML verlangt werden würden. Ganz allgemein könnte man sagen, der Konverter sollte beliebige Text-Formate ineinander umwandeln, spezielle Detail-Informationen innerhalb der Texte erkennen und als isolierte Felder ausweisen.



Die InfraSoft hat für die APA einen sehr flexiblen Zeitungs-Konverter entwickelt.

Erfüllt werden konnte dieser Anforderungskatalog durch eine offene Architektur, die das Vorhaben der APA langfristig absichert. Der Konverter wurde als Programm realisiert, das über zwei Hebel gesteuert werden kann. Einerseits durch Programm-Parameter, die im Wesentlichen Ausgangs- und Zielformat angeben. Damit weiß der Konverter, ob es sich bei der Inputdatei um eine XML-Datei, eine SGML-Datei oder ein sehr spezifisches Format einer bestimmten Zeitung handelt. Über die Programm-Parameter erfährt er ebenfalls, welches Format er erzeugen soll. Den zweiten Hebel, mit dem die Konversionen gesteuert werden, bieten individuelle Konfigurationsdateien. Über diese Konfigurationsdateien wird sehr genau beschrieben, wie jede Zeitung

das jeweilige Format nützt. So kommen zum Beispiel die Daten mehrerer Zeitungen im XML-Format, einzelne Felder wie Überschrift, Rubrik- oder Ressortangaben eines Artikels haben aber immer unterschiedliche Längen und Positionen. Und genau diese Informationen findet der Konverter in der Konfigurationsdatei, die für jede einzelne Zeitung individuell angelegt wird. Das stellt auch sicher, dass der Konverter sehr gut zu warten ist. Treten kurzfristige Änderungen in den Formaten auf, so sind die notwendigen Anpassungen fast immer über die Konfigurationsdateien möglich. Der Konverter selbst muss nur in den seltensten Fällen erweitert und neu compiliert werden.

Das gewählte Design hat sich in der Praxis sehr bewährt. Das war bereits während der Entwicklung festzustellen, denn die Format-Informationen der rund 100 Zeitungen standen während der Entwicklung erst nach und nach zur Verfügung. Und immer war es möglich, auch die neuen Daten-Quellen in den Konversionsvorgang einzubinden.

InfraSoft und Hewlett-Packard

Der von der InfraSoft entwickelte Konverter geht zur Zeit bei der APA in Produktion. Gerhard Hausberger, Projektleiter vom Generalunternehmer *Hewlett-Packard* meint zur Zusammenarbeit mit der InfraSoft:

InfraSoft als Partner

bietet Ihnen professionelle Unterstützung.

- **Wir setzen modernste Methoden ein**
Unsere Mitarbeiter verfügen über fundierte Kenntnisse in C++, Java, SQL, Notes und HTML/XML.
- **Wir stellen Know-how bereit**
Die InfraSoft-Teams haben langjährige Erfahrungen mit Datenbanken, Networking und Internet.
- **Wir schonen Ihr Budget**
Gearbeitet wird *standby* - sie bezahlen uns nur dann, wenn Sie uns brauchen.
- **Wir übernehmen Verantwortung**
Unsere Teams sind bereit, komplexe Probleme in Eigenregie zu lösen.

"Das Konvertierungsprojekt war Teil der kompletten Neugestaltung der APA Nachrichtenverarbeitung. Bei so wichtigen Projekten können wir natürlich nur fachlich kompetente Partner einsetzen, wie wir sie in den Spezialisten der InfraSoft gefunden haben. Die reibungslose Zusammenarbeit und die termingerechte Realisierung der Konvertierungslösung bestärken uns in der Auffassung, auch bei zukünftigen Projekten wieder die Spezialisten der InfraSoft einzubinden."

Die Profis von InfraSoft sind interessant für mich. Bitte kontaktieren Sie mich:

Herrn/Frau: _____
Firma: _____
Adresse: _____
Tel/DW: _____
Datum: _____

Fax-Feedback 1 505 08 06 -28 