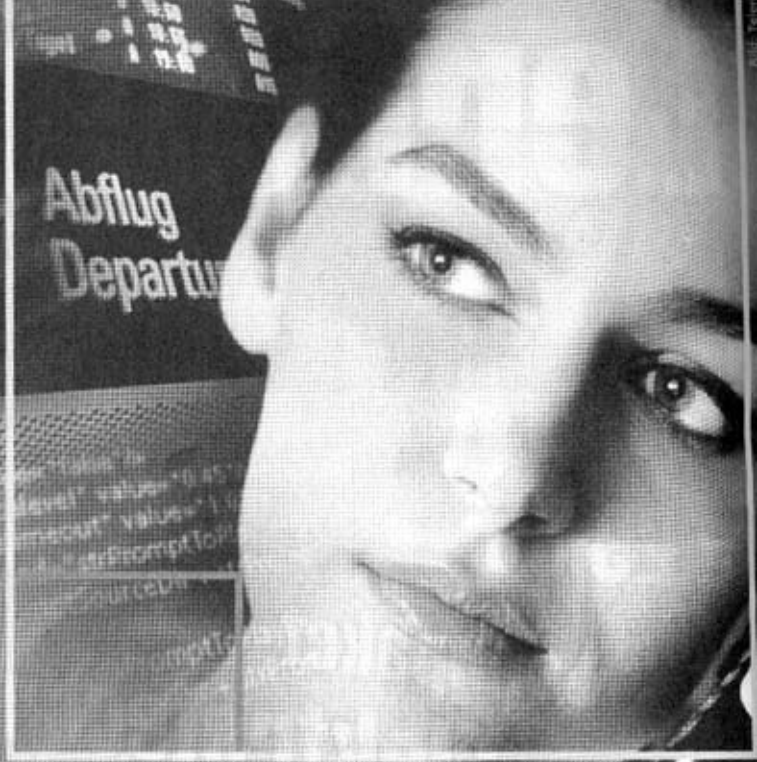
CallCenter World® 2003



Standards für Sprachanwendungen

Von Christoph Kartes

Besonders im Callcenter oder bei Infodiensten bieten Sprach-Dialogsysteme enormes Einsparpotenzial. Zur Erstellung solcher Lösungen stehen die Auszeichnungssprachen VoiceXML und SALT zur Verfügung.

Zu jeder Zeit und an jedem Ort die neusten Informationen: das World Wide Web macht es möglich. Mit so genannten Voice-Portalen lassen sich diese Infos sogar per Sprache mit einem herkömmlichen Telefon oder Handy abrufen. Auch Callcenter haben eine große Zahl von Standardanfragen, die durch eine Sprachapplikation abgearbeitet werden können. Dazu ist ein folgendermaßen strukturiertes System nötig: Vor einem Dokumenten- oder Webserver muss ein Server mit Voice-Browser installiert sein, der die Sprachbefehle so umwandelt, das sie vom Webserver verstanden und entsprechend umgesetzt werden können. Die Kommunikation zwischen Webserver und Voice-Portal kann dabei mit zwei Beschreibungssprachen erfolgen: entweder mit VoiceXML (Extensible Markup Language) oder mit SALT (Speech Application Language Tags).

VoiceXML, der gültige Standard

VoiceXML ist eine Textauszeichnungssprache, mit der Sprachapplikationen genauso einfach wie Internetanwendungen entwickelt werden können. Der Grund: VoiceXML basiert auf XML (Extensible Markup Language). Das W3C (World Wide Web Consortium, www.w3.org) hat VoiceXML im Mai 2001 als Standard zur Entwicklung von Sprachanwendungen anerkannt. Derzeit steht die Version 2.0 als Final Draft bereit. Das VoiceXML-Forum (www.voicexmlforum.org) fördert aktiv den Gebrauch dieser Auszeichnungssprache.

VoiceXML wurde entworfen, um mit der existierenden Webinfrastruktur zusammen zu arbeiten und damit die hohen Investitionen von Unternehmen in Bezug auf Web- und Applikationsserver zu schützen. Sprachanwendungen können die Kundenbindung erhöhen und erlauben den schnellen Zugriff auf Informationen. Darüber hinaus lässt sich mit ihnen die Kundenbetreuung personalisieren, was dem Support eines Callcenters gleicht. Der Standard beschreibt Aspekte der Kommunikation zwischen Mensch und Maschine. Darunter fallen die Ausgabe von synthetischer Sprache oder Audiodateien, die Erkennung von Sprache beziehungsweise Tonwahl-Impulsen und Aufnahme von Spracheingabe sowie Telefondienste.

VoiceXML oder SALT: Die Anwendung entscheidet



Bild: Crealog

Michael Kloos, Geschäftsführer der CreaLog GmbH, München über die beiden Auszeichnungssprachen: „VoiceXML und SALT haben beide ihre Existenzberechtigung, abhängig vom jeweiligen Einsatz der konkreten Kundenanwendung. VoiceXML bietet den Vorteil einer optimierten Benutzerführung für Voice-Portale, während SALT die Benutzerführung für die visuelle und sprachliche Darstellung beim Einsatz intelligenter Endgeräte vereint. Da Crealog das Ziel hat, für die Kunden immer optimale Ergebnisse bei der Sprachdialog-Gestaltung zu erreichen, setzen wir derzeit den Schwerpunkt auf VoiceXML und auf eigene Dialog-Generatoren."

Eine typische VoiceXML-Implementation besteht aus einem VoiceXML-Übersetzer (Interpreter) und einem Übersetzer-Kontext. Der Interpreter ist für die Ausführung des Codes verantwortlich und kontrolliert in Echtzeit simultane VoiceXML-Applikationen. Der Kontext ist für Unterstützungstätigkeiten verantwortlich, wie zum Beispiel das Laden des Ausgangsdokumentes bei einem eingehenden Anruf. Außerdem ruft er den Interpreter auf, wenn ein Sprachbefehl ankommt. Beide arbeiten mit Spracherkennungs- und Text-to-Speech-Engines (TTS) sowie Media-Server zusam-

men. Diese Infrastruktur wird per Voice-Gateway an das PSTN (Public Switched Telephone Network) angeschlossen.

Die Bestandteile der VoiceXML-Plattform interagieren untereinander sowie mit dem Web Application Server um eine VoiceXML-Anwendung auszuführen. Ein typischer Ablauf könnte folgendermaßen aussehen:

- Ein Benutzer leitet einen Anruf mit analogem oder digitalem Telefon ein;
- Die Telefoninfrastruktur empfängt den Anruf und die VoiceXML-Plattform führt den passenden Dialog aus;
- Eine Ansage wird mit Hilfe des Media-Servers abgespielt (vorher aufgenommene oder per TTS generierte Audiodaten);
- Benutzereingabe (Request) erfolgt als gesprochenes Wort oder Tonwahl-Impuls. Bei Spracherkennung vergleicht das System den Input mit einer vorgegebenen Grammatik und gibt bei Übereinstimmung die gefundene Textfolge zurück;
- Die Plattform führt die Logik durch, die im VoiceXML-Dokument definiert ist;
- Die Ausgabe (Response) wird per Media-Server abgespielt (vorher aufgenommene oder per TTS generierte Audiodaten).

Der softwaretechnische Prozess nach der Benutzereingabe läuft folgendermaßen ab: Die VoiceXML-Plattform schickt den Request zum Web Application Server (WAS). Dieser schaut bei der Back-end-Infrastruktur nach, ob dort Daten oder Informationen zu der Anfrage bereit liegen. Auf diesen Daten aufbauend erstellt der WAS ein dynamisches beziehungsweise nutzt ein statisches VoiceXML-Dokument. Dieses sendet er als Response zurück an die Plattform.

Da die VoiceXML-Plattform auch HTTP (Hypertext Transfer Protocol) unterstützt, kann die vorhandene Netzinfrastruktur für Sprachanwendungen verwendet werden. Entwickler, die mit verschiedenen Scripting-Technologien wie beispielsweise Perl, CGI (Common Gateway Interface) und PHP (Hypertext Preprocessor) vertraut sind, können ihre bevorzugte Technik weiter verwenden, um VoiceXML-Dokumente schnell zu erstellen.

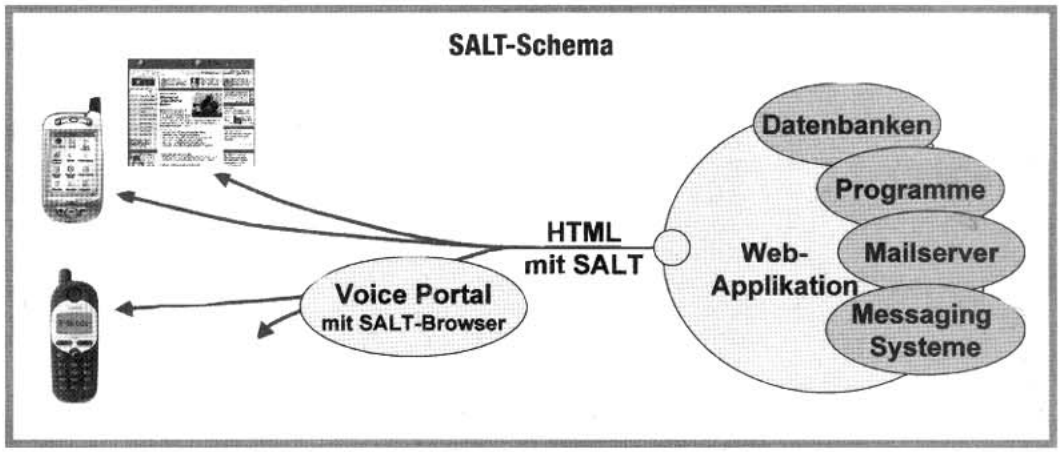
Eine neue Spezifikation: SALT

Im Oktober 2001 haben die Unternehmen Cisco, Comverse, Intel, Microsoft, Philips und Speechworks das SALT-Forum (www.saltforum.org) gegründet, um eine Auszeichnungssprache (SALT, Speech Application Language Tags) für multimodale und telefonische Sprachapplikationen zu entwickeln. Multimodal bedeutet dabei, dass beispielsweise eine Fluganfrage per Sprache erfolgt und verschiedene Flugzeiten auf dem Display eines PDA dargestellt werden. Der Anwender wählt mit dem Stift die gewünschte Route aus und bestätigt die Buchung per Stimme. Das Forum hat derzeit über 50 Mitglieder. Die Spezifikation besteht aus einen Set von XML-Elementen, welche die Integration von Sprach-Interfaces in eine Webseite ermöglichen. SALT kann mit HTML, XHTML und anderen Standards genutzt werden, um Schnittstellen für reine Sprach- und multimodale Applikationen zu schreiben.

Das SALT-Forum hat letztes Jahr im Sommer die Version 1.0 vorgestellt und sie anschließend dem W3C vorgelegt. In dem Draft zur VoiceXML Version 2.0 ist SALT noch nicht enthalten, da es etwas zu spät eingereicht wurde. In der Nachfolgeversion wird es aber integriert sein. In der SALT-Spezifikation 1.0 sind die drei Hauptbereiche Sprachein- und -ausgabe sowie die Verbindungssteuerung abgedeckt. Der Prompt-Tag erlaubt Applikationen synthetische Sprache und Audiodaten auszugeben. Listen- und Bind-Tags unterstützen die Spracherkennung durch das Verarbeiten von gesprochenem Input.

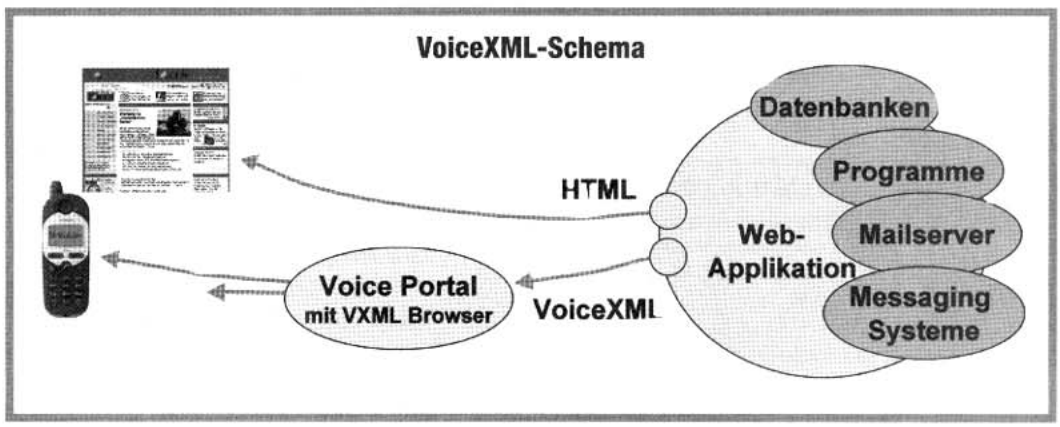
Damit lassen sich Sprachapplikationen erstellen, die direkt auf der Internetseite integriert sind. Das SALT-Forum arbeitet derzeit an einer Nachfolgeversion. Das W3C geht davon aus, dass es im zweiten Quartal 2003 den ersten Draft des multimodalen SALT vorgelegt bekommt.

SALT soll so einfach wie möglich sein. Dafür wurde die Spezifikation nach bestimmten Grundsätzen entwickelt, etwa dass SALT die einfache Integration von Sprache in Webseiten ermöglicht oder existierende Standards für Grammatik, Sprachausgabe und Semantik verwendet. Ein weiterer wichtiger Punkt ist die Trennung der Sprachschnittstelle von den Daten. Denn SALT erweitert keine vorhandene Programmiersprache sondern es stellt das Sprachinterface als eigenen Layer dar, der auf unterschiedlichen Auszeichnungssprachen aufgesetzt werden kann. Der Dialograhmen, der die Voice-Schnittstelle steuert, kann so lose oder fest wie nötig mit der zu Grunde liegenden Datenstruktur (zum Beispiel einem HTML-Formular) verbunden werden. Dadurch lassen sich Sprach- und Dialog-Komponenten auf mehreren Seiten beziehungsweise Anwendungen verwenden.



Quelle: Crealog

Die SALT-Anwendung wird im Gegensatz zu VoiceXML wie ein zusätzlicher Layer über die vorhandenen Schnittstellen gelegt: in diesem Fall die HTML-Schnittstelle. So können auch mehrere Anwendungen die gleiche Sprach-Applikation nutzen



Quelle: Crealog

Die VoiceXML-Anwendung wird wie eine zusätzliche Schnittstelle an die Webapplikation angebunden

Unterschiede von SALT und VoiceXML

VoiceXML und SALT sind beides Auszeichnungssprachen, die eine Sprachschnittstelle beschreiben. Jedoch arbeiten sie dabei sehr unterschiedlich. VoiceXML ist für Telefonie-Anwendungen bestimmt. Es wurde entwickelt, um IVR-Applikationen (Interactive

CallCenter World® 2003

Größeres VoIP-Wachstum durch SALT


Bild: Cisco

Carsten Queisser, Produkt Marketing Manager bei Cisco Systems sieht die beiden Sprachen als Ergänzung zueinander: „Die Möglichkeit, grafische Anwendungen um Sprache zu erweitern, wird den VoIP-Markt (Voice over Internet Protocol) weiter ankurbeln. SALT ermöglicht die innovative und dialogfähige Kommunikation über ein Single-Packet-Netzwerk. Endanwender können die Sprach-, Text-, oder Grafikschnittstellen von SALT-basierenden Applikationen unabhängig oder zusammen nutzen. Es ist also nicht mehr erforderlich, für jeden Ausgabemodus eine eigene Applikation zu erstellen. Cisco entwickelt SALT und VoiceXML weiter, da sich beide Sprachen ergänzen. Für SALT erwarten wir eine W3C-Spezifikation für dieses Jahr."

Unterstützung beider Standards am sinnvollsten


Bild: Voiceobjects

Tiemo Winterkamp, Head of Product Strategy & Research, VoiceObjects AG über Sprachverarbeitung: „Voiceobjects verfolgt den Ansatz der dynamischen Markup-Generierung zur Anrufzeit und kann so unterschiedlichen Ansätzen wie VoiceXML oder SALT gerecht werden. Auch entwickeln sich Standards im Sprachverarbeitungsmarkt rapide, da war die reine Fokussierung auf nur eine Sprache für uns nicht sinnvoll. Zukünftig werden Sprachdialogsysteme, insbesondere im mobilen Bereich, stärker mit Multimediaelementen kombiniert werden und auch die direkte Interaktion mit den Modalitäten, also Grafik und Sprache, wird eine stärkere Rolle spielen. Voiceobjects ist mit seiner Markup-Engine dafür gerüstet."

Voice Response) die Vorteile des World Wide Web zugänglich zu machen. VoiceXML ist eine einfache Auszeichnungssprache, die das Schreiben von systemgesteuerten Sprachdialogen über Telefone erleichtert.

SALT zielt auf Sprachapplikationen in einem ganzen Spektrum an Endgeräten ab, einschließlich Telefon, PDA, Tablet-PC und Desktop-PC. Da viele Geräte auch Displays besitzen, sind multimodale Interaktionen ein Hauptfokus. Entwickler nutzen SALT mit vorhandenen Programmier-Standards um system- und benutzergesteuerte Sprachdialoge sowie multimodale Anwendungen zu erstellen.

Darüber hinaus enthält VoiceXML viele Tags, die ein Daten- und Ausführungsmodell zusätzlich zur Sprachschnittstelle definieren. Es arbeitet nicht nur mit der Benutzerschnittstelle (<prompt>, <grammar>), sondern auch mit Daten-Modellen (<form>, <field>) und Prozeduren (<if>, <then>, <goto>). Dies wurde entwickelt, um eine allein stehende Auszeichnungssprache zur Beschreibung typischer Telefoniedialoge zur Verfügung zu stellen.

SALT hat nur wenige Elemente, weil es sich auf die Sprachschnittstelle konzentriert (zum Beispiel <prompt> und <listen>). Da SALT entworfen wurde, um auf unterschiedlichen Endgeräten zu arbeiten, läuft es innerhalb verschiedener Web-basierter Anwendungsumgebungen. So definiert es nicht ein Ausführungsmodell an sich, nutzt jedoch vorhandene Modelle wie HTML (Hypertext Markup Language), WML (Wireless Markup Language) und SMIL (Synchronized Multimedia Integration Language) mit dem Document Object Model (DOM). Dadurch können Applikationen außerhalb existierender Anwendungen erstellt werden, die einen multimodalen Dialog auf verschiedenen Endgeräten erlauben.

Ein weiteres Unterscheidungsmerkmal sind die verwendeten Standards: Beide nutzen Standards des W3C. VoiceXML setzt als Grammatik- und Sprachausgabe-Format SRGS (Speech Recognition Grammar Specification) und SSML (Speech Synthesis Markup Language) ein. SALT nutzt diese ebenfalls, darüber hinaus jedoch noch NLSML (Natural Language Semantics Markup Language) für die Spracherkennung und CCXML (Call Control Extensible Markup Language) für die telefonische Verbindungssteuerung. Diese vier Spezifikationen liegen derzeit beim W3C als Working Drafts vor.

Ausblick und Fazit

Obwohl SALT und VoiceXML entwickelt wurden, um unterschiedliche Probleme zu lösen, konvergieren diese Bereiche bereits. Die Initiativen beider Gruppen zeigen in dieselbe Richtung. So gehen Experten davon aus, dass es in Zukunft nur noch einen einheitlichen Standard geben wird, der die beiden Bereiche Telefonie und Multimodalität gemeinsam abdeckt.

So haben beispielsweise IBM, Opera und Motorola eine Spezifikation namens XHTML+Voice (X+V) dem W3C vorgelegt, welche die Verwendung von VoiceXML im multimodalen Bereich beschreibt. Dabei wird großen Wert darauf gelegt, dass X+V kompatibel zu bestehenden VoiceXML-Applikationen ist. Diese Spezifikation ist dem endgültigen SALT (Telefonie- und Multimodal-Teil) gleichzusetzen und wird voraussichtlich standardisiert werden. Welcher Name dieser Standard dann erhält, steht derzeit noch nicht fest.

In der jetzigen Situation müssen sich die Hersteller von Sprachportalen jedoch entscheiden, welche Spezifikation sie in ihre Produkte einbauen. Dabei ist es mehr eine Frage des Geschmacks als der technologischen Überlegenheit. Denn SALT in der jetzigen Version 1.0 kann VoiceXML fast gleichgestellt werden, da nur die telefonischen Funktionen enthalten sind. Einige Hersteller gehen einen zweigleisigen Weg und implementieren bereits beide Sprachen in ihre Lösungen. Somit überlassen sie dem Kunden die Entscheidung, welche Technik er einsetzen möchte. *(CK)*

Die Zukunft ist multimodal


Bild: IBM

Wolfgang Karbstein, Manager Voice Systems & Industry Solutions bei IBM Central Region, über die Chancen der Spracherkennung: „VoiceXML hat in den letzten zwei Jahren eine Menge Impulse bekommen. Nach Ansicht der Analytiker gibt es mehr als 50.000 VoiceXML-Entwickler alleine in den USA. Jetzt, wo die Welt in Richtung einer multimodalen Zukunft geht, bildet VoiceXML zusammen mit XHTML die Grundlage einer neuen Spezifikation. Beim W3C haben IBM, Oper und Motorola X+V (XHTML+Voice) eingereicht. VoiceXML entwickelt sich außerdem in Richtung natürlicher Spracherkennung. Dadurch können Nutzer ganz normal mit Sprachdialogen kommunizieren. VoiceXML findet großen Anklang bei Entwicklern, weil damit leicht Applikationen für Callcenter, Voice-Portale, Auskunft oder Sprachwahl auf Unternehmensniveau geschrieben werden können. Zusätzlich ist VoiceXML der Entwicklergemeinschaft vertraut und basiert auf offenen Standards."