



Seminario

Web multimodale: le interfacce vocali e gli standard XML

Corso di
Elaborazione del Linguaggio Naturale
Chiar.mo Prof. A.Cappelli

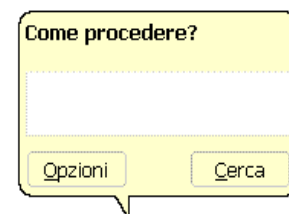
A.A. 2007-2008
A cura di Alice Guidi

Obiettivo

- Necessità di interfacce che permettono i più svariati accessi: pen-screen, voce..
- Usare il parlato per navigare e interagire con il web
- Vantaggi per i diversamente abili
- Evoluzione: integrazione con i linguaggi di markup

Sommario

- Lo scenario e il background
- Riconoscimento del parlato, piattaforme multimodali
- Piattaforme vocali: linguaggi e specifiche standard rilasciate dal W3C per il multimodale
- Esempi concreti di interazione vocale
- Riepilogo e conclusioni
- Riferimenti



Scenario

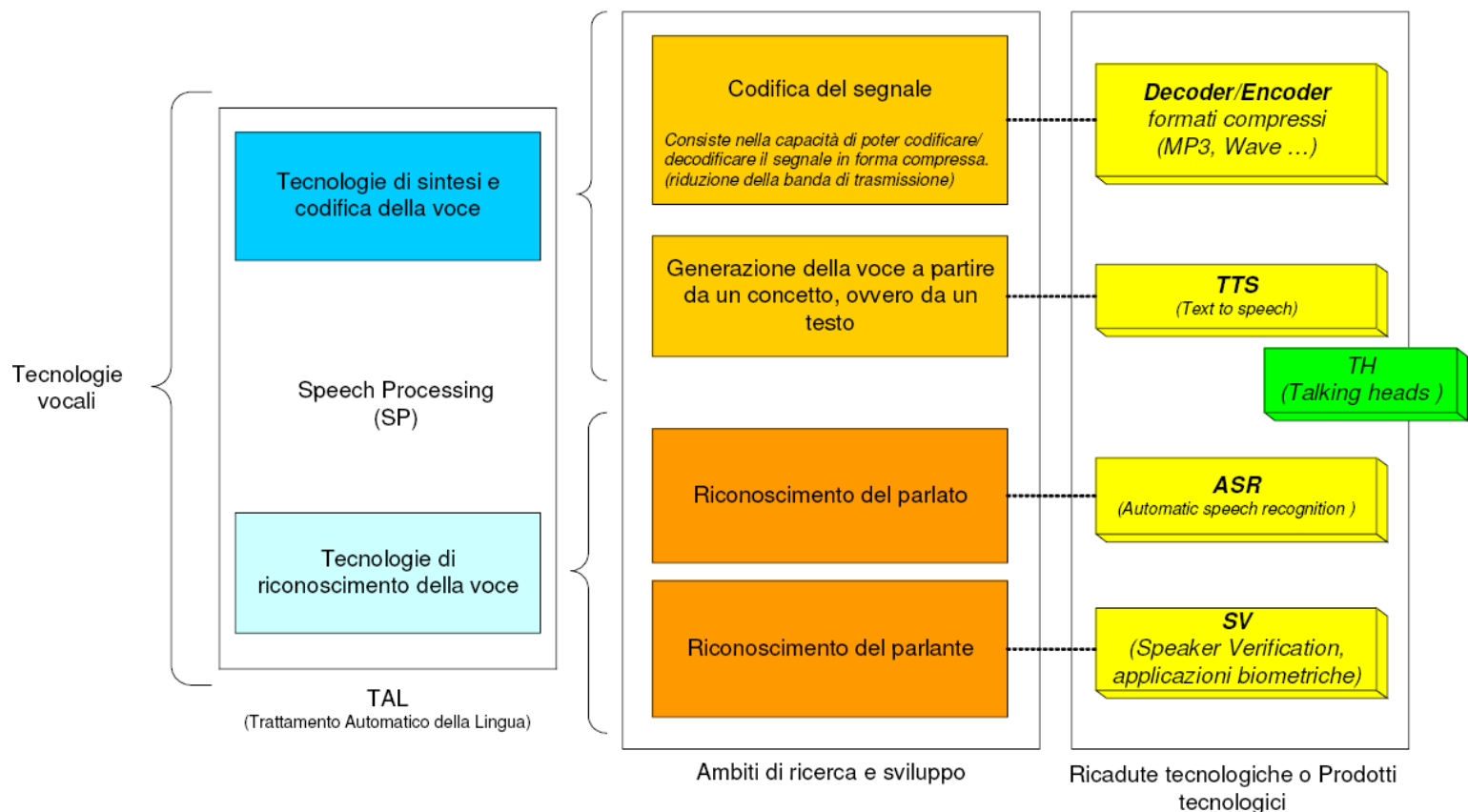
- Negli ultimi anni si è sentita maggiormente la necessità di poter accedere a diverse interfacce che permettessero i più svariati accessi: pen-screen, interazione vocale...
- Tutto ciò si trasforma in vantaggio per i diversamente abili.
- Per l'interfaccia vocale serve: un linguaggio per la programmazione, un browser multimodale, un motore TTS (Text To Speech Recognition) che trasforma la voce in testo, ASP(Automatic Speech Recognition) che confronta l'input con delle grammatiche prestabilite che a loro volta trasformano la voce in testo.
- Vedremo tutte le specifiche rilasciate dal W3C, standard basati sul linguaggio XML.



Background

Inquadramento generale

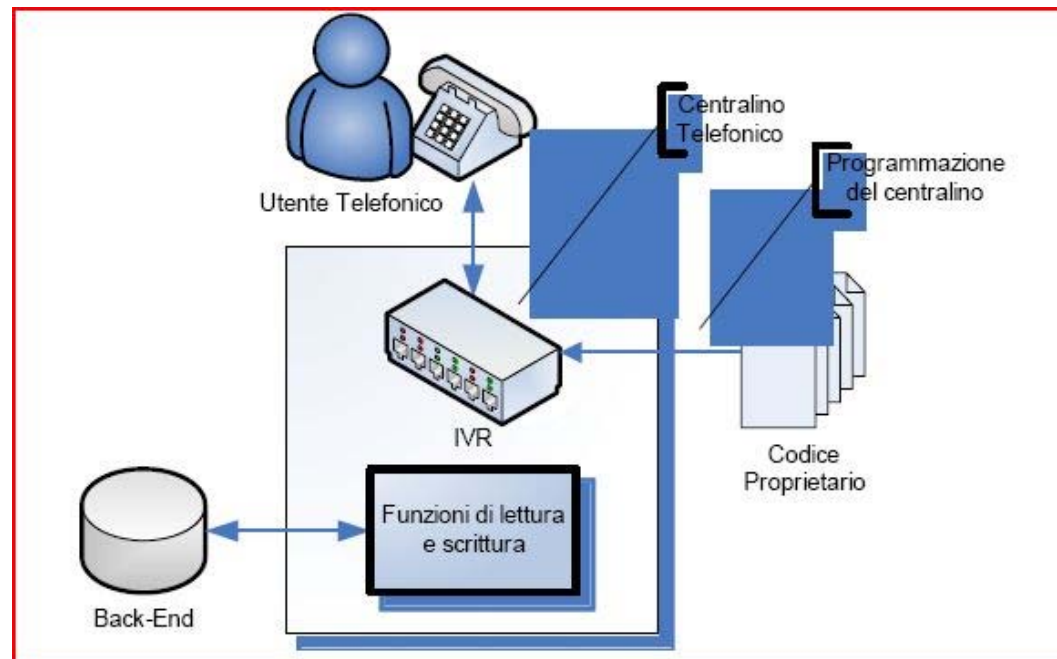
- Fino alla metà degli anni '90, non sono stati raggiunti buoni risultati nella creazione di linguaggi sintetici che suonassero naturali (provenivano da una registrazione). Negli ultimi anni si stanno evolvendo sempre più i sistemi di riconoscimento del parlato: vi è stata una proliferazione di meccanismi TTS e ASR. Contesto applicativo:



Insieme di tecniche per dialoghi automatici

- **IVR** (Interactive Voice Response), ossia risposta vocale interattiva;
- **ASR** (Automatic Speech Recognition), consente il riconoscimento vocale, assegnando a una forma d'onda acustica una o più parole;
- **IVR** (Interactive Voice Response): la tecnologia di supporto alla corretta gestione del dialogo.

Canali di comunicazione dell' IVR





Evoluzione

Comunicazione multimodale

- La comunicazione multimodale è intesa come la comunicazione accessibile attraverso diversi canali (audio, video, tattile). Un sistema multimodale è un sistema hardware/software che consente la ricezione, l'interpretazione, il processamento in input e la generazione in output di due o più modi interattivi in maniera integrata e coordinata.
- La multimodalità può migliorare l'*accessibilità* a servizi e informazioni.

Accessibilità

“La forza del web sta nella sua universalità. L’accesso da parte di chiunque, indipendentemente dalla disabilità, ne è un aspetto essenziale”.

Tim Berners Lee, Direttore del W3C

Requisiti per la progettazione di interfacce multimodali

- Per permettere un'interazione tra l'uomo e la macchina, quanto più possibile simile a quella uomo-uomo, sono state definite delle linee guida (raccolte in cinque macro categorie):





Interfacce vocali: linguaggi XML

Interfacce vocali

Linguaggi presentati:

- VoiceXML
- SSML
- SRGS
- SALT
- Più due specifiche: EMMA e InkML

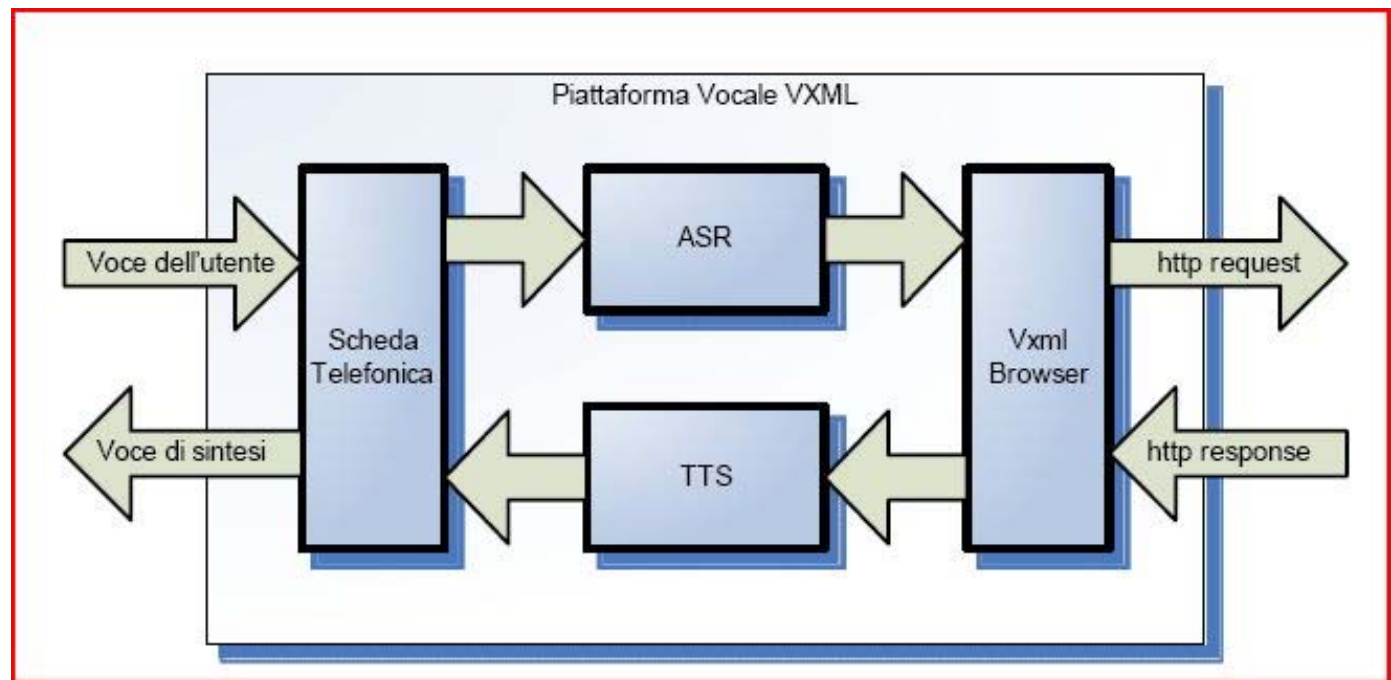


Interfacce vocali: VoiceXML

- **VoiceXML**: è stato progettato per la realizzazione di interfacce vocali. È in grado di supportare: sintetizzazione audio, audio digitalizzato, riconoscimento vocale, registrazione audio, telefonia. Uno degli scopi principali è quello di consentire la diffusione dei vantaggi dello sviluppo web e della distribuzione dei contenuti alle applicazioni vocali interattive tradizionali.
- La **versione 1.0** fu pubblicata nel Marzo del 2000. VoiceXML 1.0 venne largamente adottata, grazie anche alla sua somiglianza con HTML. **VoiceXML 2.0** è nominata *Recommendatio* W3C a Marzo 2004.
- **VoiceXML 2.1**, diventata *Recommendatio* W3C nell'Ottobre 2007, aggiunge alcune caratteristiche addizionali alla versione precedente con cui è retro-compatibile.
- Molte applicazioni commerciali supportano il VoiceXML, viene utilizzato per il tracciamento delle merci, per l'accesso vocale alle e-mail, per ascoltare le informazioni relative alla navigazione stradale.

Multimodale: piattaforme vocali

- Architettura logica di un **VoiceXML**: processo di trasformazione dalla voce in testo con markup



Altri linguaggi standard

- Il W3C ha separato la struttura dall'interfaccia vocale in altri diversi linguaggi:
- **SRGS**: *Speech Recognition Grammar Specification* è il linguaggio usato per scrivere le grammatiche (insiemi di regole) utilizzate dal riconoscitore vocale per interpretare gli input utente.
- **SISR**: *Semantic Interpretation for Speech Recognition* gli sviluppatori si avvalgono dell'interpretazione semantica, inserendo le opportune istruzioni nelle grammatiche, per manipolare i risultati ottenuti dal riconoscimento vocale in base al contesto applicativo.
- **PLS**: *Pronunciation Lexicon Specification* viene utilizzato sia dal riconoscitore vocale sia dalla sintesi per definire la pronuncia di una parola.
- **SSML**: *Speech Synthesis Markup Language* viene usato per migliorare la prosodia delle frasi per esempio aumentando o diminuendo la velocità di lettura.
- **CCXML**: *Call Control eXensible Markup Language* linguaggio per la gestione delle telefonate. Permette di accettare una chiamata in entrata, effettuare una chiamata in uscita, disconnettere una chiamata e connettere insieme più chiamate.

VoiceXML: il prologo

- Prologo
- L'elemento essenziale da conoscere per scrivere del codice VoiceXML è il **<prolog>**: ovvero la **sezione di un documento XML che contiene le dichiarazioni XML**; esso deve possedere il tag root xml, e nel nostro caso quello <vxml>, e può avere una document type declaration (DTD), commenti, e istruzioni di processing.
- Il prologo, nella forma riportata nelle specifiche W3C appare come segue:
- **Listato 1.0: Prolog Voicexml**
- ```
<?xml version="1.0" encoding="UTF-8"?>
<vxml xmlns="http://www.w3.org/2001/vxml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3.org/2001/vxml
http://www.w3.org/TR/voicexml20/vxml.xsd"
version="2.0">
```

# VoiceXML statico

- VoiceXML statico

```
<?xml version="1.0"?>
```

```
<vxml version="2.0">
```

```
 <menu>
```

```
 <prompt>Salve, il tuo genere cinematografico preferito è: commedia, giallo o
 fantascienza?
```

```
 </prompt>
```

```
 <choice dtmf="1" next="http:// ... /commedia.vxml">Commedia</choice>
```

```
 <choice dtmf="2" next="http:// ... /giallo.vxml">Giallo</choice>
```

```
 <choice dtmf="2" next="http:// ... /fantascienza.vxml">Fantascienza</choice>
```

```
 </menu>
```

```
</vxml>
```

# VoiceXML dinamico

- VoiceXML dinamico

```
<?xml version="1.0"?>
```

```
<vxml version="2.0">
```

```
<form>
```

```
 <field name="pass" type="digit">
```

```
 <property name = "inputmodes" value = "dtmf"/>
```

```
 <prompt>
```

Per visualizzare la pagina inserire la propria password

```
 </prompt>
```

```
 <filled> Il valore inserito è <value expr="pass"/>.
```

```
 <submit next="./.../getDocPassBased.asp"
```

```
 namelist = "pass"/>
```

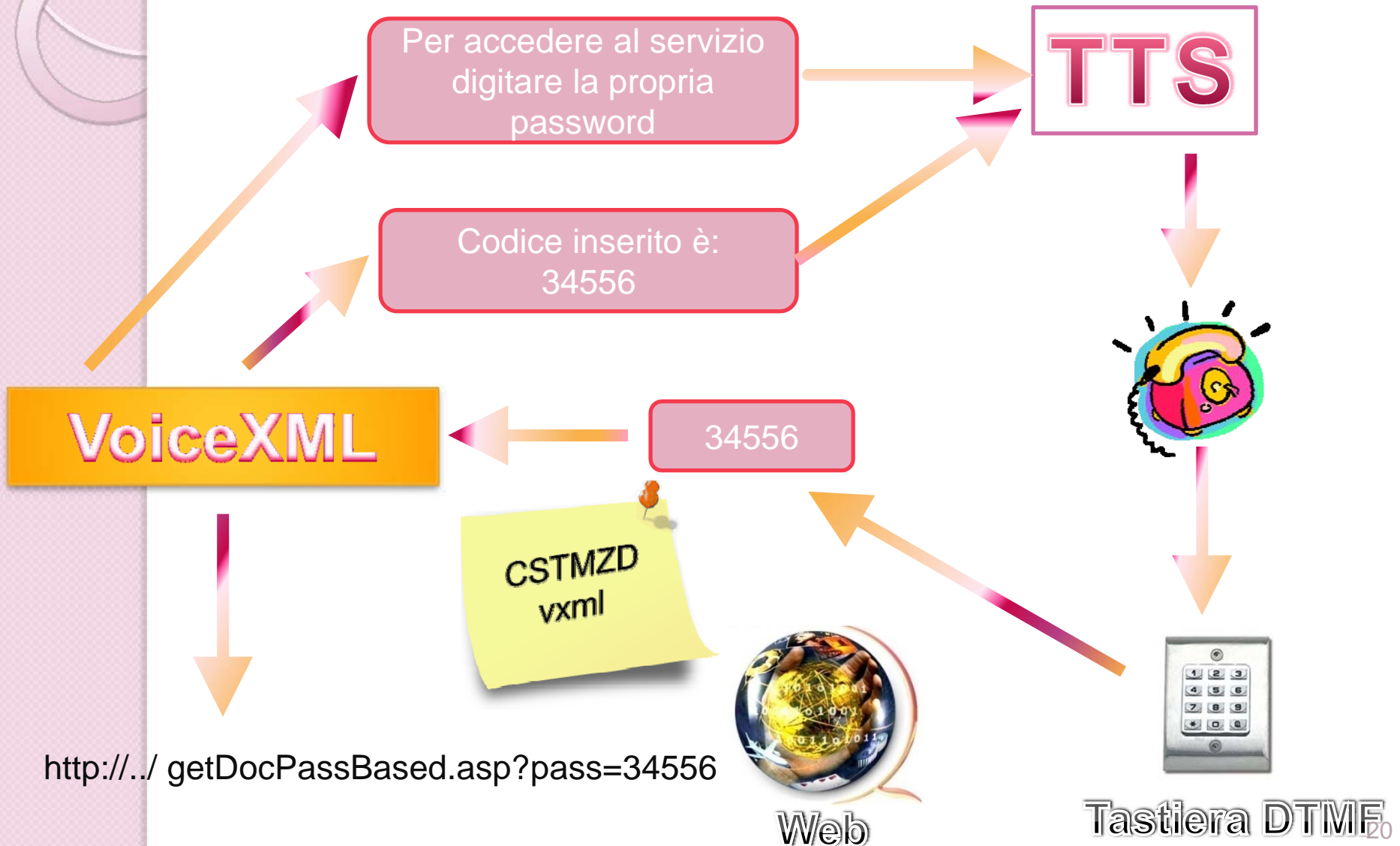
```
 </filled>
```

```
 </field>
```

```
</form>
```

```
</vxml>
```

# Esempio di esecuzione del VoiceXML dinamico



# Multimodale: XHTML+VoiceXML

## Multimodal Flight Query

Leaving From:

Arriving At:

Travel Date:

Return Date:

Type: ☐ Round Trip ☐ One Way  
☐ Multi-leg

Speech Enabled Visual Element

```
<t id="from" name="to" size="20">
 ="inputfocus" ev.handler="#voice_city"/>
```

XML Event

Time of Day:  
☐ am ☐ pm

Time of Day:  
☐ am ☐ pm

```
<vxml:form id="voice_city">
 <vxml:field name="field_city">
 <vxml:grammar src="city.jsgf" type="application/x-jsgf"/>
 <vxml:prompt> Please enter your departure city. </vxml:prompt>
 <vxml:catch event="help nomatch noinput">
 For example, say either Chicago or O'Hare.
 </vxml:catch>
 <vxml:filled>
 <vxml:assign name="document.getElementById('from')"expr=field_city"/>
 </vxml:filled>
 </vxml:field>
</vxml:form>
```

Wholly contained VoiceXML Form

# Per testare l'applicazione: strumenti e risorse

- Per pubblicare e testare un'applicazione abbiamo bisogno di un **Voice Service Provider (VSP)**. Un VSP, dovrà fornirci:
  - uno spazio su cui pubblicare i progetti;
  - i motori di ASR e TTS in lingua italiana;
  - una numerazione sul territorio italiano;
  - e possibilmente una licenza d'uso gratuita;
- Un ultimo requisito, da non sottovalutare, è la **compatibilità con gli standard** del linguaggio VoiceXML, SSML, e SRGS.
- Tra le diverse soluzioni che soddisfano i suddetti requisiti ce ne sono due: **LoquendoCafè**, della italiana Loquendo, e **Voxbuilder** dell'azienda irlandese Voxpilot, che eroga servizi di telefonia sul nostro territorio nazionale grazie a Map Telecom, e attraverso Skype.
- In particolare, **Loquendocafe**: è lo spazio messo a disposizione da Loquendo per testare applicazioni vocali. Il servizio ha diverse caratteristiche tra cui:

un'**area tutorial** divisa in VoiceXML tutorial e Grammar Tutorial, un'**area dedicata alle reference** dei linguaggi VoiceXML e SSML, **diversi esempi di codice** già sviluppati per cominciare ad impraticchirsi con il linguaggio, e, infine, un'**area riservata agli iscritti**, dove una volta autenticati potremmo accedere alla nostra console per gestire gli upload, e testare l'applicazione. All'interno di questa sezione **LoquendoCafe mette a disposizione un debugger VoiceXML** e un tool per testare le grammatiche, due strumenti fondamentali per gli sviluppatori.

# VoiceXML user group

- È nata anche una user group intorno allo standard VoiceXML: VIUG VoiceXML Italian user group, la quale cura il sito [www.vxmitalia.it](http://www.vxmitalia.it) dando sussidio a esperti e principianti del settore che vogliono avvicinarsi a questa nuova tipologia di interazione con il web.

# Aziende che utilizzano VoiceXML

- BeVocal Café
- Conita
- Eloquent
- HP
- IBM
- Intervoice's
- JVoiceXML
- Loquendo
- ... approfondiamo Loquendo, caso italiano.



# Applicazione reale:



- Loquendo è una società del gruppo Telecom con sede a Torino ed uffici in tutto il mondo.
- Con oltre 30 anni di ricerca e di sviluppo, Loquendo si pone come leader mondiale nel mercato delle tecnologie vocali con la sua capacità di fornire piattaforme e soluzioni complete per la creazione di applicazioni e servizi vocali di ultima generazione.
- Loquendo utilizza la piattaforma **Vox Natura** che integra i due standard VoiceXML e CCXML.
- Prodotti offerti: “Loquendo TTS” (sintesi vocale), “Loquendo ASR” (riconoscimento del parlato), “Loquendo Speaker Verification” (verifica del parlatore), integrazione con piattaforme di tipo IVR (Interactive Voice Response).
- Impiego e supporto in molte aziende: servizi e-banking, automazione dei processi di customer care e contact center, vocalizzazione di pagine web.
- 25 lingue disponibili.

# Loquendo piattaforma Vox Natura

*La piattaforma VoxNauta permette alle imprese, ai gestori di telecomunicazioni, alle aziende di servizi di poter sviluppare applicazioni vocali seguendo l'architettura Web e basate sugli standard W3C VoiceXML e CCXML.*

*Il CCXML rende estremamente flessibile il controllo di chiamata, mentre il VoiceXML permette una gestione sofisticata dell'interazione vocale.*

*La piattaforma VoxNauta può integrare le tecnologie vocali Loquendo, le più avanzate attualmente presenti sul mercato, fornisce il supporto completo di tutti gli standard più importanti e offre diverse funzionalità estremamente innovative per garantire le migliori prestazioni nelle applicazioni vocali.*

## Supporto completo degli standard



# SSML e SRGS

- **SSML** (Speech Synthesis Markup Language) e **SGRS** (Speech Recognition Grammar Specification) sono due standard rilasciati dal W3C che interoperano con VoiceXML.
- SSML fa parte di un grande gruppo di specifiche di markup per i Voice Browser sviluppati dal W3C. È progettato sullo standard XML per l'aiuto e la creazione di un linguaggio sintetico nel web e in altre applicazioni. Il ruolo principale del linguaggio SSML è quello di dare agli autori di sintesi vocale uno standard sul controllo dei vari aspetti tipici del parlato come: **la pronuncia, il volume, il tono, la velocità** ecc.
- Il markup può essere prodotto automaticamente per esempio con l'aiuto di XSLT o CSS3 (con un documento XTHML) oppure può essere sviluppato manualmente da un esperto.

# SSML: criteri di design

Consistency

Internationalization

Interoperability

Generation  
and  
Readability

Generality

Implementable

# SSML: codice

- Esempio del codice SSML: è presentato l'elemento <voice> il quale possiede gli attributi: **xml:lang**, **age**, **gender**, **variant**, **name**.

```
<?xml version="1.0"?>
```

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://www.w3.org/2001/10/synthesis
 http://www.w3.org/TR/speech-synthesis/synthesis.xsd" xml:lang="en-US">
```

```
 <voice gender="female">Mary had a little lamb,</voice>
```

```
 <!-- now request a different female child's voice -->
```

```
 <voice gender="female" variant="2"> Its fleece was white as snow.
</voice>
```

```
 <!-- processor-specific voice selection -->
```

```
 <voice name="Mike">I want to be like Mike.</voice>
```

```
</speak>
```

# SRGS

- **SGRS** (Speech Recognition Grammar Specification) è il linguaggio che indica la sintassi di parole o gruppi di parole che un riconoscitore vocale può aspettarsi come input.
- La sintassi della grammatica si presenta in due forme distinte: Augmented BNF (ABNF) Form e XML Form. La specifica assicura che entrambe le rappresentazioni siano semanticamente mappabili per permettere trasformazioni automatiche tra le 2 forme.
- Lo standard SGRS del W3C è modellato sulla specifica JSGF (JSpeech Grammar Format), il quale è proprietà di Sun Microsystems, Inc. California, U.S.A.

# SGRS: codice

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE grammar PUBLIC "-//W3C//DTD GRAMMAR 1.0//EN"
"http://www.w3.org/TR/speech-grammar/grammar.dtd">
<grammar xmlns="http://www.w3.org/2001/06/grammar" xml:lang="en"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://www.w3.org/2001/06/grammar
 http://www.w3.org/TR/speech-grammar/grammar.xsd" version="1.0"
 mode="voice" root="basicCdm">
 <meta name="author" content="music"/>
 <rule id="basicCdm" scope="public">
 <example> close the window </example>
 <example> delete a image </example>
 </rule>
 ...
</grammar>
```





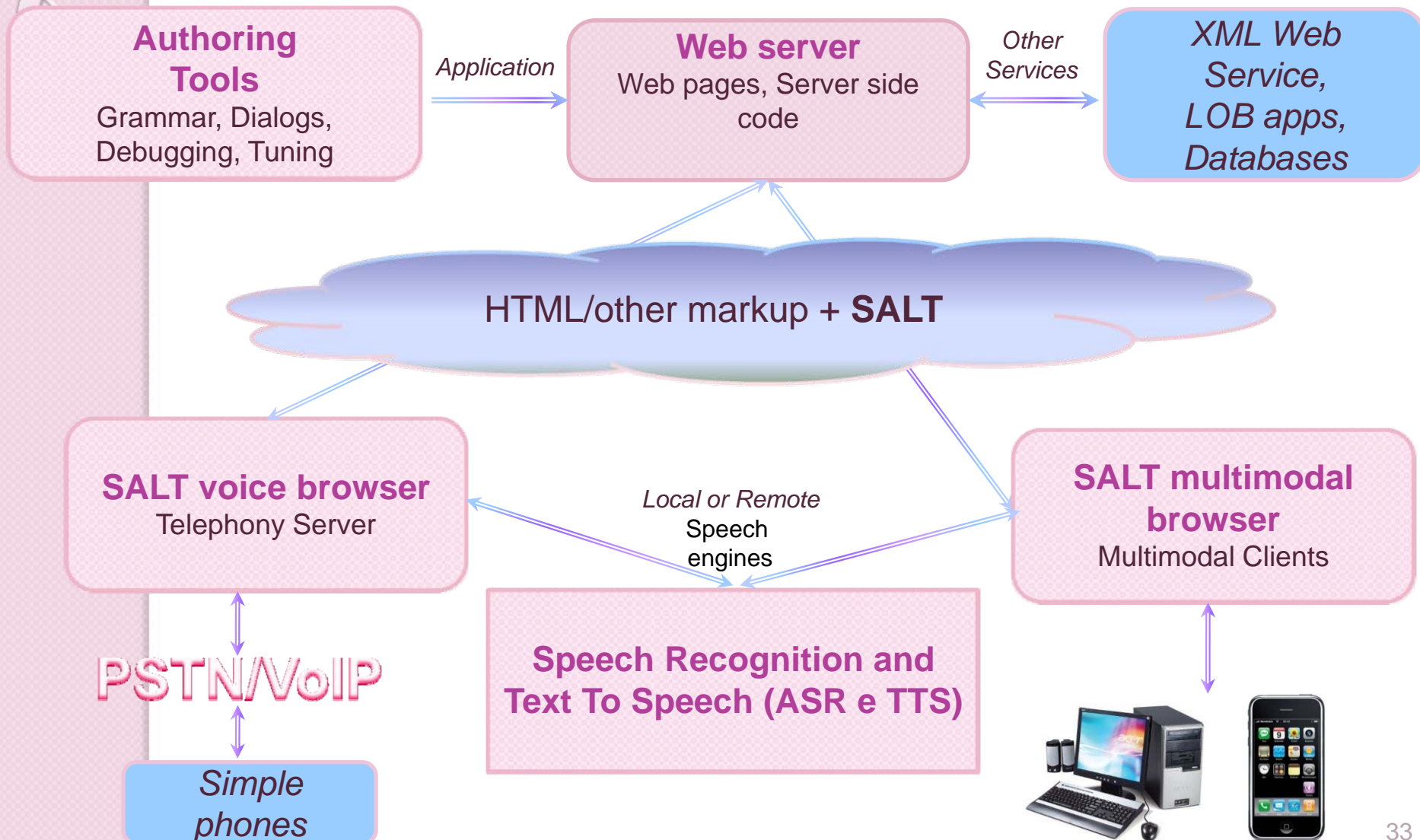
# SALT

- **Salt**, Speech Application Language Tags, è un linguaggio di markup per interfacce vocali. È costituito da un gruppo di elementi XML. Ad essi sono associati attributi e proprietà, eventi, metodi di oggetti DOM (Document Object Model), che forniscono un'interfaccia vocale alla pagina web.
- Salt 1.0 è stato pubblicato a Luglio 2002 nel forum ad esso dedicato.
- È un linguaggio royalty-free, interagisce con altre tecnologie Web standard come HTML, XHTML, SMIL.
- Riutilizza i principali standard del W3C come: SSML, EMMA, CCXML.



# Multimodale: linguaggio utilizzato

- Architettura di un sistema SALT-based



# Elementi principali di SALT

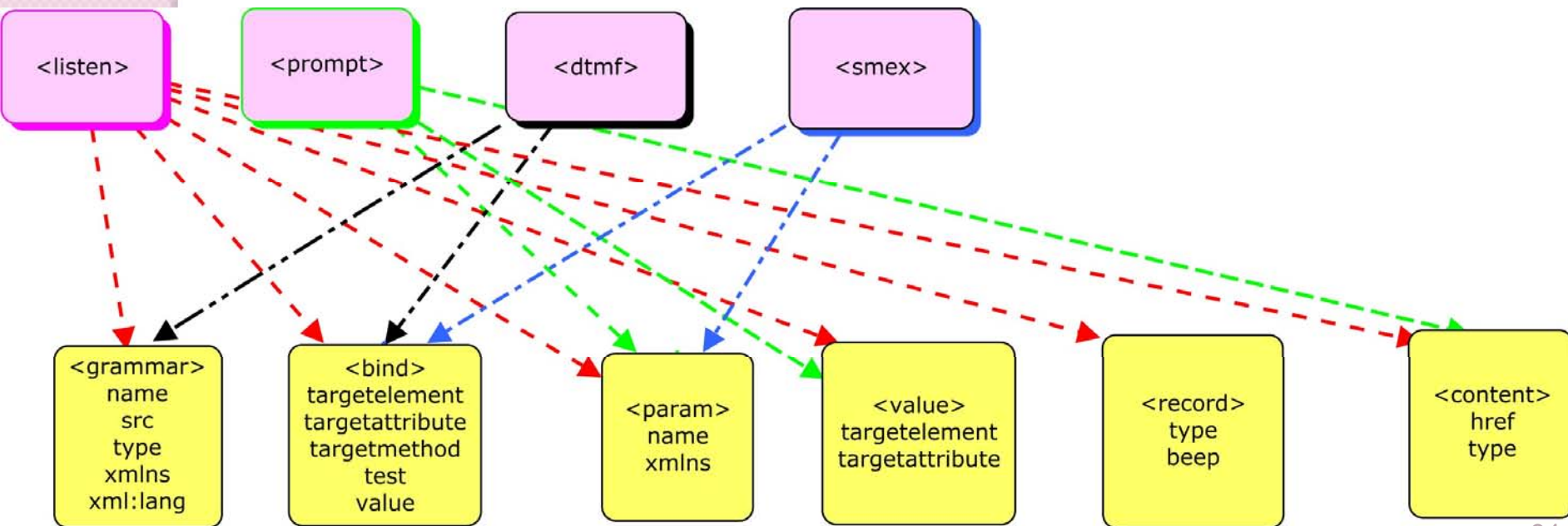
I quattro elementi principali di SALT sono:

**<listen>**: per la configurazione della sintesi del parlato;

**<prompt>**: per la configurazione del riconoscimento del parlato, post-processing e la registrazione;

**<dtmf>**: per il controllo di collezioni DTMF;

**<smex>**: per scopi generali della comunicazione con componenti della piattaforma.



# Codice SALT (1)

```
<html xmlns: salt=http://www.saltforum.org/2002/SALT>
<body onload="RunAsk()">
 <form id="travelForm">
 <input name="textBoxOriginCity" type="text" />
 <input name="textBoxDestCity" type="text" />
 </form>
<!-- Speech Application Language Tags -->
<salt:prompt id="askOriginCity"> Da dove vuoi partire? </salt:forum>
<salt:prompt id="askDestCity"> Dove vuoi andare? </salt:forum>
<salt:listen id ="recoOriginCity" onreco="procOriginCity()">
 <salt:grammar src="city.xml" />
<salt:listen>
<salt:listen id ="recoDestCity" onreco="procDestCity()">
 <salt:grammar src="city.xml" />
<salt:listen>
<!-- scripted dialog flow -->
```

*continua..*

# Codice SALT (2)

```
<script>
function RunAsk() {
 if (travelForm.txtBoxOriginCity.value=="") {
 askOriginCity.Start();
 recoOriginCity.Start(); }
 else if (travelForm.txtBoxDestCity.value=="") {
 askDestCity.Start();
 recoDestCity.Start(); }
 }

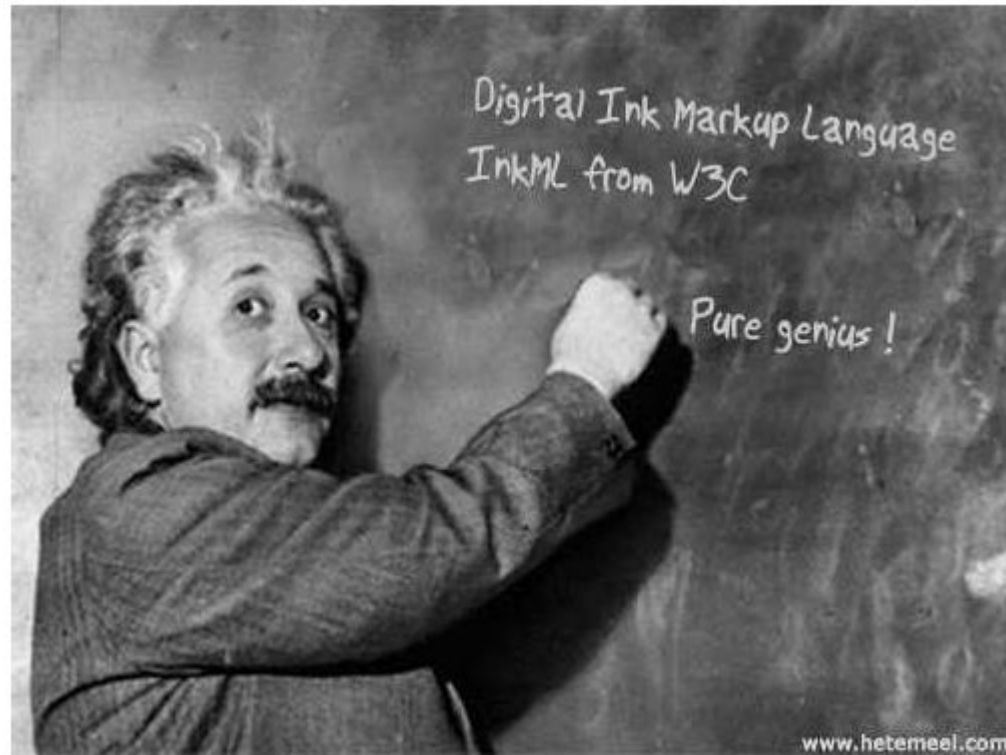
function procOriginCity() {
 travelForm.txtBoxOriginCity.value = recoOriginCity.text;
 RunAsk(); }
function procDestCity() () {
 travelForm.txtBoxDestCity.value = recoDestCity.text;
 travelForm.submit(); }
</script>
</body>
</html>
```

# Salt e html

- Come modificare una pagina html per supportare il Salt

```
1. <html xmlns:salt="http://www.saltforum.org/2002/SALT">
2. <head>
3. <title>My First Multimodal Application</title>
4. </head>
5. <body onload="RunIt()">
6. <h3>This is my first Multimodal application!</h3>
7. <salt:prompt id="first">
8. This is my first Multimodal application!
9. </salt:prompt>
10. </body>
11. <script language="javascript">
12. function RunIt() {
13. first.Start();
14. }
15. </script>
16. </html>
```

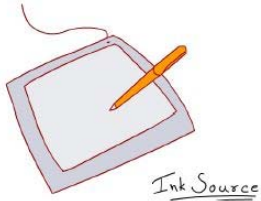
# Nuove specifiche





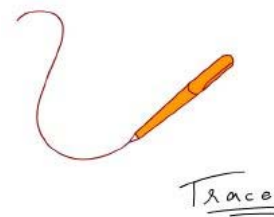
# Emma e InkML

- **Multimodale interaction framework:**  
rappresenta a livello astratto la struttura di un'architettura che si basa sull'interazione multimodale con gli utenti e descrive quali sono i linguaggi di markup che reputa più conformi ai vari componenti di un'ipotetica applicazione.
- In particolare vedremo: **Emma** e **InkML**  
(altri linguaggi per le interfacce vocali come VoiceXML, SSML, SALT sono già stati trattati).



# InkML

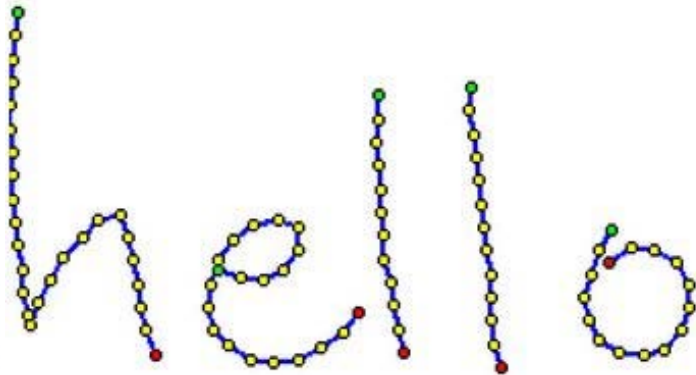
- Sempre più spesso sono disponibili devices con pen interfaces utili per manipolare le informazioni. La scrittura è un modo per l'input familiare per molti utenti. La mancanza pertanto di un formato standard per "l'inchiostro digitale" ha severamente limitato la cattura, la trasmissione, il processamento e la presentazione dei dati tra devices eterogenei.
- In risposta a ciò nasce **InkML**, Ink Markup Language, il quale fornisce un semplice formato di dati per promuovere l'interscambio di inchiostro digitale tra diverse applicazioni.
- InkML stabilisce un set di elementi primitivi sufficienti per l'utilizzo in applicazioni base.
- Tutto il contenuto di un documento InkML è racchiuso nel tag `<ink>`. Il dato fondamentale in un file InkML è l'elemento `<trace>`, il quale rappresenta una sequenza contigua di punti: sono date ad es. le coordinate x e y che individuano la posizione della penna. Una sequenza di tracce sono accumulate in unità di senso (caratteri o parole).





# InkML: esempio di pen input

- Input grafico con la penna, sono evidenziate le tracce

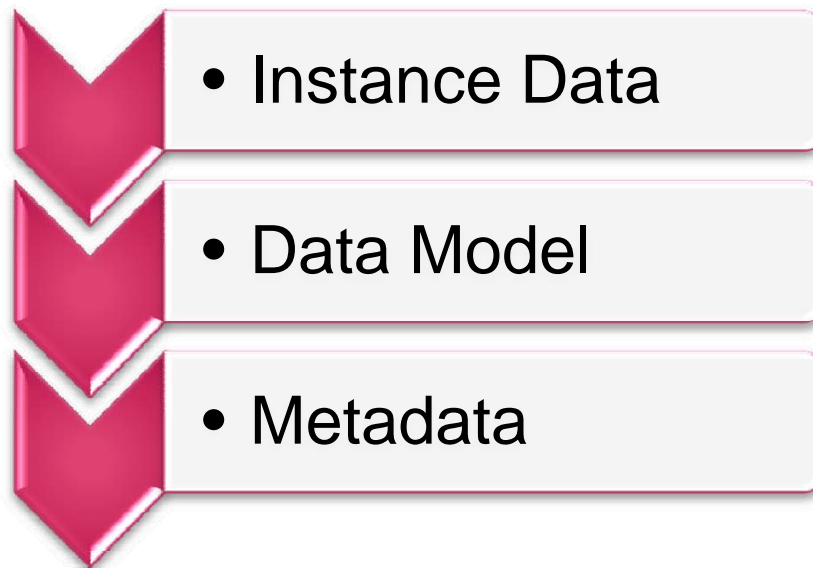


- Il tracciamento per la scritta *hello*

```
<ink>
 <trace>
 10 0 9 14 8 28 7 42 6 56 6 70 8 84 8 98 8 112 9 126 10 140
 13 154 14 168 17 182 18 188 23 174 30 160 38 147 49 135
 58 124 72 121 77 135 80 149 82 163 84 177 87 191 93 205
 </trace>
 <trace>
 130 155 144 159 158 160 170 154 179 143 179 129 166 125
 152 128 140 136 131 149 126 163 124 177 128 190 137 200
 150 208 163 210 178 208 192 201 205 192 214 180
 </trace>
 <trace>
 227 50 226 64 225 78 227 92 228 106 228 120 229 134
 230 148 234 162 235 176 238 190 241 204
 </trace>
 <trace>
 282 45 281 59 284 73 285 87 287 101 288 115 290 129
 291 143 294 157 294 171 294 185 296 199 300 213
 </trace>
 <trace>
 366 130 359 143 354 157 349 171 352 185 359 197
 371 204 385 205 398 202 408 191 413 177 413 163
 405 150 392 143 378 141 365 150
 </trace>
</ink>
```

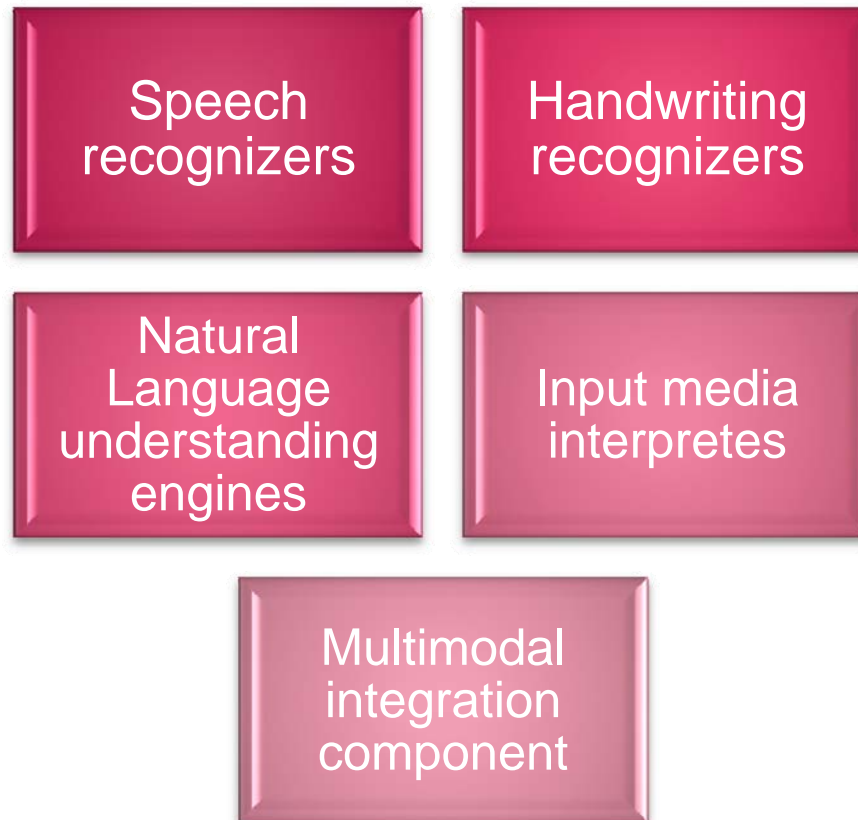
# EMMA

- EMMA è stato programmato per l'uso in sistemi che forniscono una interpretazione semantica per una varietà di input (natural language text, GUI, ink input).
- EMMA si focalizza sull'annotazione e l'interpretazione dell'informazione di input singoli e composti.
- EMMA è ancora in fase di sviluppo, ma ha già grandi potenzialità per integraZIONE DEI diversi devices sul Web.
- Un documento tipico di EMMA è composto da tre parti:



# Componenti di EMMA

- Componenti che generano il markup di EMMA

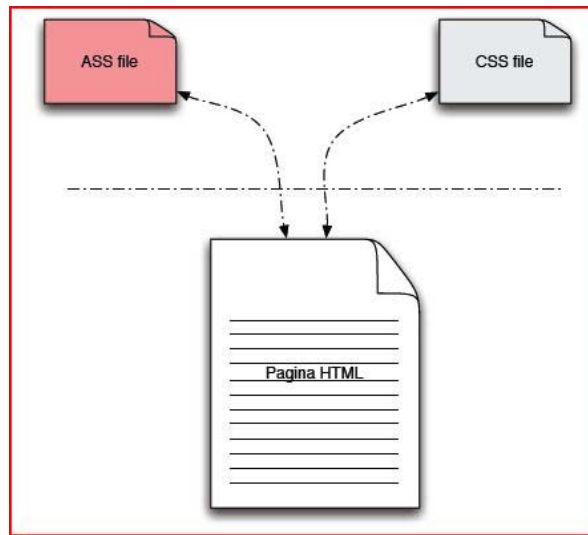




# **Presentazione dei contenuti**

# Multimodale: presentazione dei contenuti


- ASS Audio Style Sheet, estensione dei CSS per presentare i contenuti web.



La relazione tra design usato attualmente e il design ASS-enabled.

# Multimodale: presentazione dei contenuti

- Livelli espressivi supportati da ASS



The screenshot shows the ANSA.it website in a web browser. The browser's address bar displays 'http://www.ansa.it/'. The website's layout includes a top navigation bar with links like 'Notizie', 'Programmi TV', 'uni', and 'webmail ps'. A sidebar on the left contains a 'HOME' section with links to 'CANALI', 'NEWS ALERT - RSS', 'ANSA ABBONATI', 'ANSA SUL TELEFONINO', and 'ANSA INTERNATIONAL'. Below these are 'NEWS' and 'Top News' sections with links to 'News in English', 'Italia', 'Mondo', 'Sport', 'Calcio', 'Economia e Finanza', 'Cultura e Società', 'Scienza', 'Internet', 'Spettacolo', 'Moda', 'Musica', 'Cinema', and 'Regioni'. A map of Italy is also present. The main content area features a 'News' section with several articles, each with a thumbnail image and a headline. The articles include: 'UN VOLO GOVERNATIVO PORTA I FAMILIARI DELLE VITTIME A SHARM', 'IL 2 OTTOBRE SARA' FESTA DEL NONNO, SI' DEFINITIVO A LEGGE', 'IRAQ, ASSASSINATI 12 TECNICI IRACHENI DI UN DEPURATORE', 'RAI: DIRITTI TV, GIUDICE SOSPENDE GARA E DOMANI L'UDIENZA', 'POSTE: DA OGGI RACCOMANDATE DA PC CASA, PRESTO TELEFONINO', 'SPAZIO: RIUSCITO IL LANCIO DELLO SHUTTLE, METEO FAVOREVOLE', 'GIUSTIZIA: CIAMPI HA PROMULGATO LA RIFORMA', 'ARRIVA NUOVA PROTESI DEL PENE SI IMPIANTA IN DAY HOSPITAL', and 'SCIOPERO TRENI: SULT, NON VIAGGIA OLTRE META' DEI CONVOGLI'. A 'Photonews' section at the bottom shows a row of small images. The footer of the website includes the 'in' logo and the 'ANSAPHOTO' logo with the text 'Le migliori foto dal Mondo'.

livello di pagina    livello di classe    livello di elemento





# Esempi di interazione vocale

# Google Voice Search

Google™  
GOOG-411 1-800-GOOG-411



[Scarica il video](#)

Dial from any phone

1-800-GOOG-411

(1-800-466-4411)

#### About GOOG-411

Google's new 411 service is free, fast and easy to use. Give it a try now and see how simple it is to find and connect with local businesses for free.

[Learn more](#) - [FAQ](#)

Liked the video? Want to comment or guess who the voice of GOOG-411 is? Post your opinion on our [YouTube page](#).

**1** Dial 1-800-GOOG-411 from any phone



**2** State the location and business type



**3** Connect to the business for free



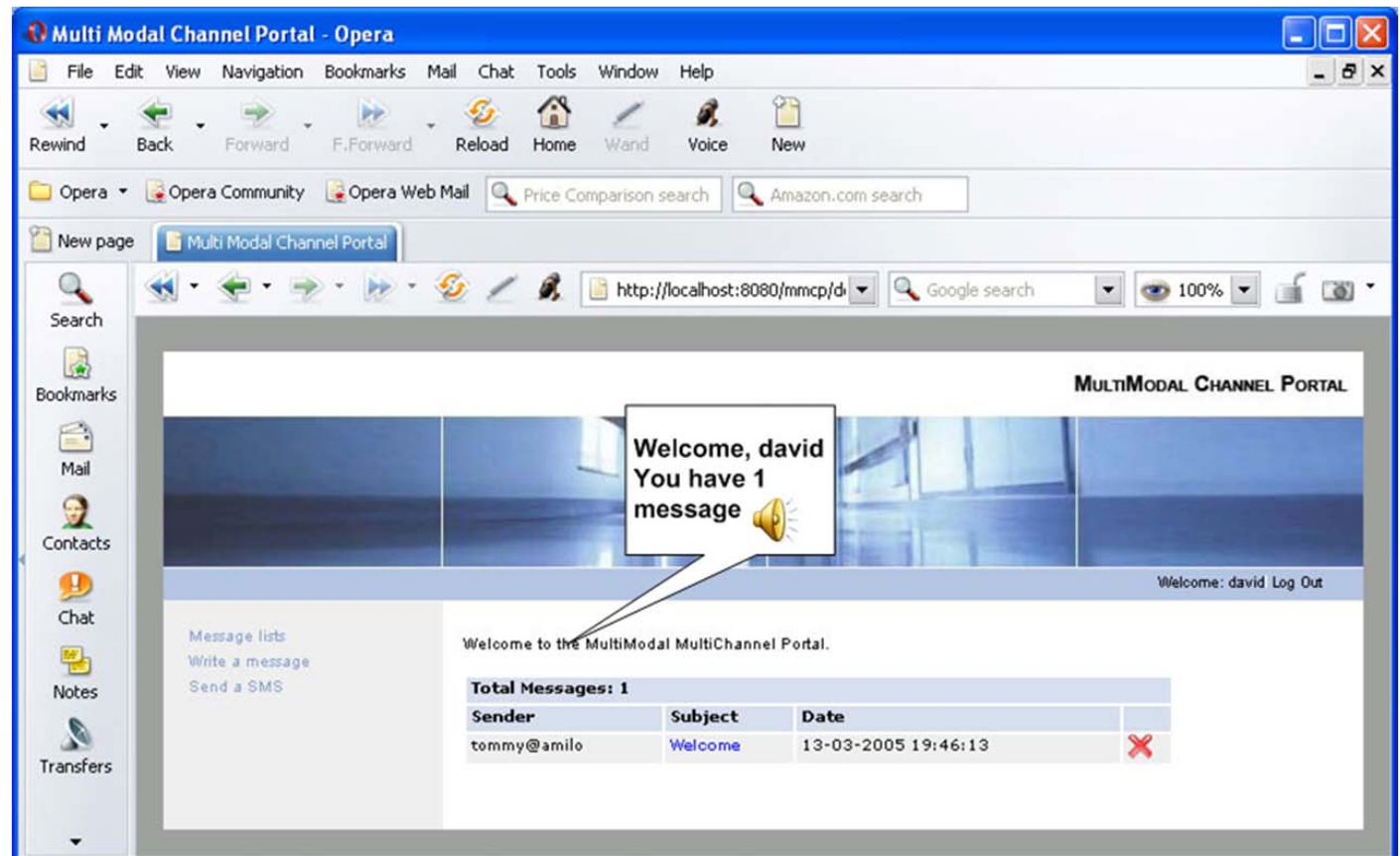
**4** Done!





# Browser Opera

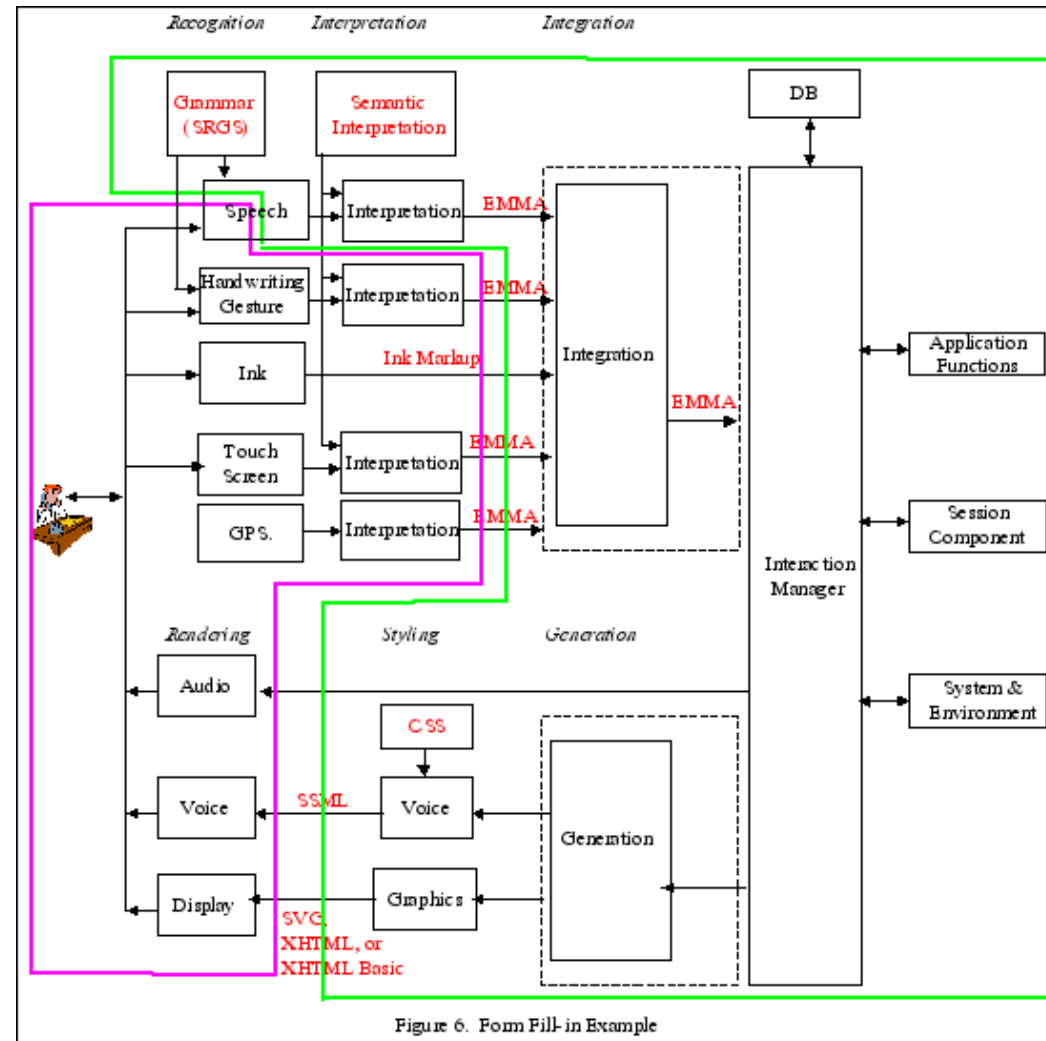
## MultiModal Channel Portal





**Riepilogando...**

# Multimodal Interaction Framework



# Multimodal Interaction Framework

- Descrizione della figura precedente.
- L'utente vuole prenotare un volo con il suo cellulare. La conversazione tra l'utente e l'applicazione si concentrerà su domande del tipo “partenza”, “destinazione”, “data di partenza e data di ritorno”.
- Area rosa: componenti allocati dal lato client, area verde componenti presenti sul server. Il riconoscimento del parlato è distribuito tra le due aree.
- L'informazione prodotta dall'utente sarà codificata con EMMA combinandola con componenti ad essa integrati.
- L'informazione dell'utente sarà così riconosciuta e interpretata per poi essere indirizzata direttamente alla interazione di controllo che ne permetterà l'integrazione con gli altri componenti.

# In conclusione...perché sviluppare applicazioni vocali?

I vantaggi raggiungibili da un' applicazione vocale sono in particolar modo:

- **accesso ai dati in modo naturale** (voce è il metodo più semplice e naturale per accedere ai nostri dati, fare domande e ricevere risposte);
- **aumenta il bacino di utenti** (non bisogna sottovalutare la diffusione della telefonia mobile);
- **accessibilità 24/24 ore** (no limiti di orario e fluidità nelle chiamate, diminuisce tempo di attesa);
- **accesso indipendentemente dallo spazio e dal tempo**;
- **portatori di handicap** (aumento delle possibilità d'impiego per portatori di handicap come ipovedenti o persone con difficoltà di movimento);
- **maggiore accessibilità** (per chi ha già le mani impegnate per il proprio lavoro);
- **tastiere piccole o assenti** (molti supporti hanno una tastiera limitata, piccola o assente).



# Riferimenti

# Sitografia

## Applicazioni, linguaggi e specifiche standard

- EMMA: <http://www.w3.org/TR/emma/>
- InkML: <http://www.w3.org/TR/InkML/>
- Loquendo: <http://www.loquendo.it/it/index.htm>
- Google Voice Search: <http://www.google.com/goog411/>
- SSML: <http://www.w3.org/TR/speech-synthesis/>
- SALT: <http://www.xml.com/pub/a/2003/05/14/salt.html>
- SRGS: <http://www.w3.org/TR/speech-grammar/>
- Voice XML: <http://www.w3.org/TR/voicexml20/>
- Voice Browser: <http://www.w3.org/Voice/>
- Requisiti multimodali: <http://www.w3.org/2002/mmi/>
- XML.com: Multimodal Interaction on the web: <http://www.xml.com/pub/a/2004/01/21/multimodal.html>
- VIUG, <http://www.vxmlitalia.com/>
- Web and XML Glossary: <http://dret.net/glossary>



# Bibliografia ragionata

## Comunicazione multimodale e interazione uomo-macchina

- Leah M. Reeves, Jennifer Lai, James A. Larson, Sharon Oviatt, T.S. Balaji, Stéfanie Buisine, Penny Colligs, Phil Cohen, Ben Karaal, e altri, *Guidelines per multimodal user interface design*. Communication of the ACM – Special Issue on Multimodal Interface, vol. 47(1), 2004.
- Samy Bengio, Hervé Bourlard (Eds.): *Machine Learning for Multimodal Interaction, First International Workshop, MLMI 2004*, Martigny, Switzerland, June 21-23, 2004.
- Harry Bunt, Robbert-Jan Beun (Eds.): *Cooperative Multimodal Communication, Second International Conference*, CMC'98, Tilburg, The Netherlands, January 28-30, 1998.



# Bibliografia ragionata

- Guideline for multimodal user interface design:  
<http://stephanie.buisine.free.fr/publis/CACM04.pdf>
- ETSI: T.Barbieri, A.Bianchi, L.Sbattella, *Multimodal communication for vision and hearing impairments*, set. 2004,  
[http://portal.etsi.org/docbox/EC\\_Files/EC\\_Files/eg\\_202191v010101p.pdf](http://portal.etsi.org/docbox/EC_Files/EC_Files/eg_202191v010101p.pdf)
- ETSI: European Telecommunications Standard Institute. Human Factors (HF); *Multimodal interaction, communication and navigation guidelines*.
- T.V. Raman, *User interface principle for multimodal interaction*, CHI 2003.
- T.V Raman, Gerald McCobb, Rafah A.Hosn. *XHTML+Voice 1.1 Versatile Multimodal Solutions*. XML Journal, apr.2003.

# Bibliografia ragionata

- Gunther Kress, Theo van Leeuwen *Multimodal discourse : the modes and media of contemporary communication*, London, Arnold 2001.
- Isabella Poggi, *Mind, hands, face and body : a goal and belief view of multimodal communication*, Berlin 2007.
- Eleonora Bilotta, *Problemi di percezione e comprensione nelle interfacce grafiche*, Edizione Preliminare, Settembre 1996, disponibile on-line al link: <http://galileo.cincom.unical.it/Pubblicazioni/editoria/libri/Comprend/coper.html>
- Paolo Baggia, *Standard per applicazioni multimodali* <http://www.vxmlitalia.com/Baggia.pdf>
- Loquendo: Paolo Baggia, *Accessibilità in voce: il linguaggio VoiceXML*, Roma 20 Giugno 2003.

# Bibliografia ragionata

- Giuseppe Baldari, Rosario Barbera, *Applicazioni multimodali basate su XML*, dispensa reperibile all'url: <http://elite.polito.it/courses/tamm/tesine/VoiceXML-dispensa.pdf>