

Laurea Specialistica in Informatica
a.a. 2007-2008

Interazione Uomo-Macchina II:

Interfacce Intelligenti

Fiorella de Rosis

Laboratori di
Irene Mazzotta e Nicole Novielli

1

Perché abbiamo bisogno di interfacce intelligenti?

- Perché le interfacce dei sistemi attuali sono troppo 'rigide'
- Perché non cambiano quando cambiano le nostre esigenze
- Perché il paradigma d'interazione attuale non è intuitivo e naturale
- ...

Cosa rende 'intelligente' una interfaccia?

- La capacità di adattarsi alle esigenze di utenti diversi
- La capacità di interfacciarsi con l'utenza nei luoghi e nei contesti più diversi
- La capacità di 'prendere l'iniziativa' e dare suggerimenti all'utente
- La capacità di interagire in linguaggio naturale con gli utenti
- La capacità di spiegare e giustificare le ragioni delle sue azioni
- ...

Henry Lieberman, MIT Boston

2

"The next generation of interfaces, often called 'intelligent', will provide a number of additional benefits to users, including *adaptivity, context sensitivity and task assistance*. As with traditional interfaces, principled intelligent interfaces should be learnable, usable, and transparent. In contrast, however, intelligent user interfaces promise to provide **additional benefits** to users that can enhance interaction, such as:

- *Comprehension* of possibly imprecise, ambiguous and/or partial multimodal input;
- *Generation* of coordinated, cohesive and coherent multimodal presentations;
- Semi or fully-automated completion of *delegated tasks*;
- Management of the interaction ... by *representing reasoning*, and exploiting *models of the user*, domain, task, and context."

M Maybury, MITRE Corporation, USA
W Wahlster, University of Saarbruecken, Germany
Readings in Intelligent User Interfaces, 1998

Esaminiamo meglio questi punti:

1. *Comprehension* of possibly imprecise, ambiguous and/or partial multimodal input;
2. *Generation* of coordinated, cohesive and coherent multimodal presentations;
3. Semi or fully-automated completion of *delegated tasks*;
4. Management of the interaction ... by *representing reasoning*, and exploiting *models of the user*, domain, task, and context.

4

1. *Comprehension* of possibly imprecise, ambiguous and/or partial multimodal input:

esempi di quesiti

A quale informazione U è interessato, in questo momento?
Cosa sta tentando di fare? Interpretazione di azioni

U sta sorridendo: perché?
U ha sbuffato: forse è stanco?
U ha inarcato le sopracciglia: forse non mi crede?
Interpretazione di espressioni

U rifiuta di fare l'esercizio: perché?
Cosa intendeva dire, Marco, quando mi ha detto: "Mah, sarà...!"
Interpretazione di frasi pronunciate

... ..

5

Comprehension of possibly imprecise, ambiguous and/or partial multimodal input

- > *Riconoscimento di comandi* (in linguaggio naturale o in altra forma) in termini di intenzioni, obiettivi, conoscenze dell'utente,...
- > *Riconoscimento di espressioni* del viso o gesti, in termini di affaticamento, 'carico cognitivo', stato emotivo, ...

6

2. *Generation* of coordinated, cohesive and coherent multimodal presentations

Come rappresento i dati che U mi ha chiesto di elaborare: in forma grafica, tabellare o componendo diversi media?

Qualche esempio:

Ipermedia dinamici:
come presento l'informazione su un sito web (ad esempio, commerciale) in modo da rendere questa informazione utile e 'attraente' (per l'utente e/o per me)?

Persuasione:
come faccio a convincere Marco a mangiare meno hamburger?

e-learning:
come spiego allo studente il concetto di 'energia'? Quale esempio gli faccio, per aiutarlo a capire meglio?

7

3. Semi or fully-automated completion of *delegated tasks*

Principio della *delega*:

Stabilire quali compiti il sistema può/deve effettuare anche senza un comando esplicito da parte dell'utente.

Qualche esempio noto:

Aggiornamento antivirus
Controllo errori in word
Download di posta elettronica, altri?

Nuove prospettive:

L'utente definisce, in fase di set-up del sistema, *quali compiti delegare* (rivedendo scelte definite per default)

8

Semi or fully-automated completion of *delegated tasks*

Criterio fondamentale:

La delega deve sempre poter essere aggiornata o ritirata.

Un esempio di delega ritirata:

Clip di Microsoft:

<http://www.cnn.com/TECH/computing/9810/16/clipdeath.idg/>
http://www.pcanswer.com/articles/lat_officexp.htm

9

4. Management of the interaction ... by *representing reasoning* and *exploiting models of the user, domain, task, and context.*"

Dato un insieme di *conoscenze specifiche* su quello che l'utente crede, desidera, preferisce, intende fare, sul dominio e sulle funzioni da svolgere e

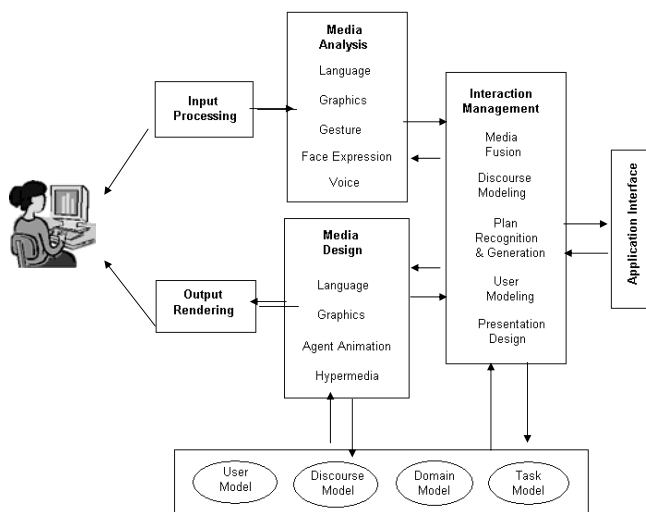
Data una *conoscenza generale* sui legami che, in genere, esistono fra questi elementi

Ragionare su questa conoscenza per inferire nuovi (probabili?) elementi dello stato mentale dell'utente, scegliere il modo migliore di effettuare il task ed eseguirlo.

Notare la differenza fra conoscenza generale e specifica, che è tipica dei 'sistemi intelligenti' e sarà ripresa più volte nel Corso.

10

Architettura di una interfaccia intelligente (da Maybury e Wahlster)



11

Cosa ci proponiamo, con questo Corso

1. Imparare a *generare un messaggio in linguaggio naturale*
2. Imparare a *riconoscere e interpretare comandi* dell'utente in linguaggio naturale
3. Imparare a *simulare dialoghi in linguaggio naturale*
4. Imparare a *tener conto delle caratteristiche dell'utente*, nello svolgere questi compiti

Vediamo subito degli esempi, che ci faranno da guida nei metodi che descriveremo durante il Corso.

12

Esempio 1: Le previsioni del tempo sul sito web di Repubblica

Un link agli approfondimenti

Cartine per due giorni

I servizi
Homepage
Previsioni
Cartine del tempo
Il tempo in Italia
Il tempo nel mondo
Copertura nuvolosa
Direzione dei venti
Mappa termica
Mari e Venti
Neve
Precipitazioni
Pressione
Satellite
Uragani
Tempo e salute
Fenomeni naturali
Quando partire
Glossario

Sei in: [Home](#) | [Il tempo in Italia](#)

OGGI ALLE ORE 12
OGGI ALLE ORE 18
DOMANI ALLE ORE 6

PROSSIMI GIORNI »

RICERCA IN ITALIA
Indica un comune

RICERCA ALL'ESTERO
Indica un paese

Prenota con Expedia
☒ Voli ☐ Hotel

Le previsioni

lunedì 26 dicembre

Nord: molto nuvoloso o coperto sulla Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e Basso Vento con precipitazioni sparse, occasionalmente nevose anche in pianura, specie nella seconda parte della giornata. Nuvolosità più attenuata sulle rimanenti zone del Nord con addensamenti più consistenti sulle zone alpine, specie sui versanti settentrionali ove potranno verificarsi delle nevicate.

Centro e Sardegna: molto nuvoloso o coperto, con piogge diffuse. Nevicate sulle zone interne a quote collinari specie su Toscana, Marche ed Umbria.

Sud e Sicilia: generalmente molto nuvoloso a tratti coperto con precipitazioni diffuse e persistenti su Campania, Calabria Tirrenica e Sicilia settentrionale. Neve sui rilievi interni oltre gli 800-1.000 metri di quota. Molto nuvoloso ma con fenomeni isolati sul resto del Sud, pur con tendenza al peggioramento.

Temperature: in generale diminuzione, più marcata sulle regioni settentrionali.

Venti: deboli o moderati settentrionali al Nord; moderati da Ovest-Sud Ovest al Centro ed al Sud, con rinforzi sulla Sardegna, sulle coste joniche e sul Salento.

Mari: da poco mossi a mossi il Mar Ligure e l'Alto Adriatico; mossi o molto mossi tutti gli altri mari.

Testo coerente con la mappa



Le previsioni

lunedì 26 dicembre

Nord: molto nuvoloso o coperto sulla Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e Basso Vento con precipitazioni sparse, occasionalmente nevose anche in pianura, specie nella seconda parte della giornata. Nuvolosità più attenuata sulle rimanenti zone del Nord con addensamenti più consistenti sulle zone alpine, specie sui versanti settentrionali ove potranno verificarsi delle nevicate.

Centro e Sardegna: molto nuvoloso o coperto, con piogge diffuse. Nevicate sulle zone interne a quote collinari specie su Toscana, Marche ed Umbria. Nuvolosità e fenomeni più attenuati sulla Sardegna.

Sud e Sicilia: generalmente molto nuvoloso a tratti coperto con precipitazioni diffuse e persistenti su Campania, Calabria Tirrenica e Sicilia settentrionale. Neve sui rilievi interni oltre gli 800-1.000 metri di quota. Molto nuvoloso ma con fenomeni isolati sul resto del Sud, pur con tendenza al peggioramento.

Temperature: in generale diminuzione, più marcata sulle regioni settentrionali.

Venti: deboli o moderati settentrionali al Nord; moderati da Ovest-Sud Ovest al Centro ed al Sud, con rinforzi sulla Sardegna, sulle coste joniche e sul Salento.

Mari: da poco mossi a mossi il Mar Ligure e l'Alto Adriatico; mossi o molto mossi tutti gli altri mari.

Stessa struttura in giorni diversi

martedì 27 dicembre

Nord: coperto con precipitazioni diffuse a prevalente carattere nevoso sulla Pianura Padana, sui rilievi liguri e sul Basso Piemonte. Nuvoloso o molto nuvoloso sulle zone alpine, ma con scarsa probabilità di precipitazioni.

Centro e Sardegna: generalmente molto nuvoloso con precipitazioni sparse, specie su Toscana, Umbria e Marche. Le precipitazioni assumeranno carattere nevoso sui rilievi interni anche a quote collinari.

Sud e Sicilia: molto nuvoloso o coperto con piogge diffuse, più consistenti sulle regioni del basso versante tirrenico specie nella seconda parte della giornata. Neve sui rilievi oltre gli 800 metri di quota.

Temperature: in ulteriore diminuzione, specie al Centro-Nord.

Venti: deboli nord-orientali al Nord, moderati occidentali o sud-occidentali al Centro-Sud, in intensificazione dal pomeriggio.

Mari: tutti mossi o molto mossi, localmente agitati il Canale di Sardegna, lo Stretto di Sicilia e lo Jonio.

Esempio 2: Dialoghi in linguaggio naturale

Vedremo:

dialoghi di *information-seeking* ad iniziativa del sistema
 dialoghi di *information-seeking* ad iniziativa mista
 dialoghi di *'advice-giving'*

Es 2.1: Un dialogo molto semplice

S: Welcome to Irene's travel agency. I'm here to help you to buy your travel tickets. Which place do you want to go?

U: London

S: Which place do you start from?

U: Bari

S: Do you prefer travelling with a regular or a low-cost company?

U: Regular

S: Which tariff do you prefer? Economic or business?

.....

17

Es 2.2: Un dialogo più 'naturale'

S: Welcome to Irene's travel agency. I'm here to help you to organize your travels.

U: I would like to go to London starting from Bari in the month of May, and have not much money to spend.

S: Oh, London! I would love going there! There are several low-cost companies traveling from Bari to London. Let's see what we can do.

Do you mind to transfer from Fiumicino to Ciampino, when you are in Rome?

U: No problem: I'm young and used to travel.

S: Good! Then, just tell me when exactly you would like to leave.

.....

18

S: A at American Express may I help you? ... i dialoghi naturali possono
U: yeah this is B. I need to make some travel arrangements essere più complessi!!
[recording notification, charge number, travel advance, phone]

S: ok and what do you need to do?

U: ok on June sixth from San Jose to Denver, United

S: leaving at what time?

U: I believe there's one leaving at eleven o'clock in the morning

S: yeah that's United flight four seventy

U: ok

S: leaves at eleven a.m. and arrives Denver at two twenty p.m. out of San Jose

U: that's the one

S: and a return?

U: return June seventh

S: leaving at what time?

U: uh I believe there's one at five sixteen

S: yeah that's United flight six fifty nine it arrives back into San Jose at six forty two

U: ok

S: let me just see what sort of fare I can get you on this

U: all right

S: ok the best fare I can get you on this would be five sixty eight round trip. That does have a twenty five percent change cancellation penalty on it

U: ok

S: And we'd have to ticket it by Tuesday

U: right, that's fine

[hotel, car, seating, mileage plus number]

U: ok thank you very much

S: thanks a lot

Es 2.3

(Trascritto di un dialogo
telefonico American
Express)¹⁹

Es 2.4: Un dialogo di 'advice-giving'

S: Dovresti andare a correre, Giuseppe!

U: Perché?

S: Perché sei giovane, ci tieni alla tua salute, e correre fa bene alla salute.

U: Ma ti pare che, a vent'anni, già devo pensare alla salute?

S: Secondo me sì. Ma comunque, correre aiuta anche a tenersi in forma.

U: E chi te l'ha detto?

S: Lo dicono studi epidemiologici svolti in diversi paesi, da istituti di ricerca qualificati.

U: Ma io detesto correre.

...

20

Oz: Ciao, il mio nome e' Valentina. Sono qui per darti dei suggerimenti su come migliorare la tua dieta.

Subject1: piacere, mi chiamo francesco

Oz: Quali sono le tue abitudini alimentari?

Subject2: tipica dieta mediterranea

...

Oz: Che cosa hai mangiato per colazione?

Subject4: latte e caffè con biscotti secchi

Oz: Che cosa hai mangiato per pranzo?

Subject5: peperoni ripieni

...

Oz: Quanta frutta mangi al giorno, in genere?

Subject7: sempre dopo i pasti

Oz: Ti piacciono frutta e verdura?

Subject8: soprattutto la frutta

...

Oz: Il tuo peso ti sembra giusto, o vorresti cambiarlo?

Subject11: devo perdere alcuni chili... cosa mi consigli?

Oz: Bisognerebbe mangiare solo se si ha fame, ed aspettare che la digestione sia completa, prima di assumere altri cibi.

Subject13: insomma, devo rispettare gli orari dei pasti

Oz: Mangiare ad orari fissi aiuta ad evitare di saltare i pasti e di trascurare sostanze che spesso non vengono compensate nei pasti successivi.

Subject14: sono d'accordo con te

...

Oz: Limitare la dose di grassi, in particolare quelli saturi, rappresenta un elemento fondamentale di una dieta sana.

Subject18: certo

Oz: Hai mai riflettuto su quanto sia importante, per te, mangiare bene e sentirti in forma?

Subject19: sì, infatti ho seguito una dieta dimagrante proprio per sentirmi in forma, perdendo quasi 40 Kg

...

Es 2.5: Un dialogo con un 'Agente Animato'



In Es 2.5: Fase 1: il Sistema avvia il dialogo

Oz: Ciao, il mio nome e' Valentina.

Si presenta

Sono qui per darti dei suggerimenti su come migliorare la tua dieta.

Descrive brevemente il suo ruolo

Subject1: piacere, mi chiamo Francesco

22

Fase 2: il Sistema cerca di capire quali sono le caratteristiche dell'Utente

Oz: Quali sono le tue abitudini alimentari?

Subject2: tipica dieta mediterranea

...

Oz: Che cosa hai mangiato per colazione?

Subject4: latte e caffè con biscotti secchi

Oz: Che cosa hai mangiato per pranzo?

Subject5: peperoni ripieni

...

Oz: Quanta frutta mangi al giorno, in genere?

Subject7: sempre dopo i pasti

Oz: Ti piacciono frutta e verdura?

Subject8: soprattutto la frutta

...

Oz: Il tuo peso ti sembra giusto, o vorresti cambiarlo?

Subject11: devo perdere alcuni chili... cosa mi consigli?

Fa domande...

... finché l'Utente non prende l'iniziativa

23

Fase 3: il Sistema dà consigli all'Utente

Oz: Bisognerebbe mangiare solo se si ha fame, ed aspettare che la digestione sia completa, prima di assumere altri cibi.

Subject13: insomma, devo rispettare gli orari dei pasti

Oz: Mangiare ad orari fissi aiuta ad evitare di saltare i pasti e di trascurare sostanze che spesso non vengono compensate nei pasti successivi.

Subject14: sono d'accordo con te

...

Oz: Limitare la dose di grassi, in particolare quelli saturi, rappresenta un elemento fondamentale di una dieta sana.

Subject18: certo

... finché l'Utente non fa obiezioni...
(ad esempio: 'Ma io adoro il pane e burro e i dolci fritti!')

24

Fase 4: il Sistema raccoglie nuove informazioni

Oz: Hai mai riflettuto su quanto sia importante, per te, mangiare bene e sentirti in forma?

Subject19: sì,
infatti ho seguito una dieta dimagrante
proprio per sentirmi in forma,
perdendo quasi 40 Kg
....

... finché l'Utente
non chiede di concludere il dialogo

25

Quali sono i principali settori applicativi delle interfacce intelligenti?

- Gestione automatica di dialoghi speech-based nei call center
- Sistemi vocali di supporto alla guida
- Generazione automatica di siti web commerciali
- Sistemi di e-learning
- Robotica
- ...

26

Chi lavora nel campo delle interfacce intelligenti?

Gruppi di ricerca industriali e universitari:

- diverse Telecom europee; AT&T
- diverse case automobilistiche (es Toyota)
- Microsoft
- MIT
- molte Università americane ed europee

Diversi software possono essere utilizzati, per sviluppare applicazioni

27

Software per il NLP

Da programmi che eseguono task elementari:

- analisi RST di testi
- generazione di frasi in LN,
- text-to-speech synthesis,
- parsing,
- riconoscimento di speech,
- simulazione di dialoghi semplici speech-based

a programmi che assemblano moduli in tool più complessi (per la simulazione di dialoghi).

Alcuni commerciali, altri free. Vediamone alcuni.

28

Siti nei quali vengono elencati i principali software disponibili

(ma anche i 'corpora' pubblici e gli articoli più significativi):

Natural Language Software Registry

<http://registry.dfki.de>

Software tools for NLP

http://www-a2k.is.tokushima-u.ac.jp/member/kita/NLP/nlp_tools.html

NLP/Information Retrieval Software Repository

<http://www.comp.nus.edu.sg/~rpnlpir/>

Building NLG Systems

www.csd.abdn.ac.uk/ereiter/buildingfaq.html

29

Software per l'analisi RST di testi

RST Tool (free)

Il sito ufficiale sulla teoria:

<http://www.sfu.ca/rst/>

Un tool originalmente costruito da Mick O'Donnell:

<http://www.wagsoft.com/RSTTool/>

In versione modificata da Daniel Marcu:

<http://www.isi.edu/licensed-sw/RSTTool/>

Rasta

RASTA (Rhetorical Structure Theory Analyzer), is a system for automatic discourse analysis

It identifies rhetorical relations present in written discourse by examining information available in syntactic and logical form analyses.

Since there is a many-to-many relationship between rhetorical relations and elements of linguistic form, RASTA identifies relations by the convergence of a number of pieces of evidence, many of which would be insufficient in isolation to reliably identify a relation.

30

Software per il NLG

RealPro

www.cogentex.com/technology/demo.shtml

FUF/SURGE: Ben Gurion University, Israel (free)

FUF: Functional Unification Formalism Interpreter

SURGE: A Syntactic Realization Grammar for Text Generation

<http://www.cs.bgu.ac.il/surge/index.html>

HALogen

www.isi.edu/publications/licensed-sw/halogen/index.html

Grammatica in Prolog della New York University

www.nyu.edu/pags/linguistics/anlebk.html

31

RealPro (CogenTex Inc)

A text generation "engine" that performs syntactic realization — i.e., the transformation of *abstract syntactic specifications* of natural language sentences (or phrases) into their corresponding *surface forms*.

It supports multiple languages and levels of linguistic representation, with performance suitable for real-world applications.

RealPro provides a grammar rule engine that can generate text from sophisticated, multi-level linguistic representations.

The abstraction it provides makes it easy to generate *many syntactic variants of the same semantic content* on demand — unlike with template-based approaches, where the combinatorics of generating multiple syntactic variants quickly becomes unmanageable.

32

Una demo in linea

RealPro Demo: Variations on "John loves Mary"

Generator Options	Tense	Progressive	Perfect	Question	Negation	Focus on Mary
	<input type="radio"/> Present <input type="radio"/> Past <input type="radio"/> Future	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes
Output	John loves Mary.					
Generator Options	Tense	Progressive	Perfect	Question	Negation	Focus on Mary
	<input type="radio"/> Present <input type="radio"/> Past <input type="radio"/> Future	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes
Output	John is loving Mary.					
Generator Options	Tense	Progressive	Perfect	Question	Negation	Focus on Mary
	<input type="radio"/> Present <input type="radio"/> Past <input type="radio"/> Future	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes
Output	John is not loving Mary.					
Generator Options	Tense	Progressive	Perfect	Question	Negation	Focus on Mary
	<input type="radio"/> Present <input type="radio"/> Past <input type="radio"/> Future	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Yes
Output	John will not love Mary.					

Input in RealPro

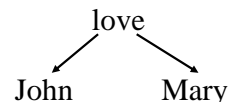
```
// -----
// John loves Mary.
// -----

DSYNTS:
love [ class:verb ]
( I John [ class:proper_noun ]
  II Mary [ class:proper_noun ]
)

END:
// -----
// John does not love Mary.
// -----

DSYNTS:
love [ class:verb tense:pres polarity:neg ]
( I John [ class:proper_noun ]
  II Mary [ class:proper_noun ]
)

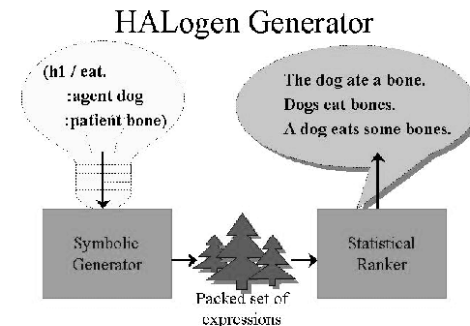
END:
```



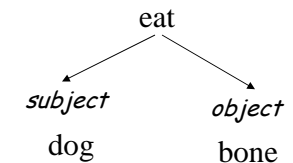
33

HALogen:

a general-purpose NLG by the Information Sciences Institute,
University of Southern California



Input: un albero sintattico con specifiche sulla forma, il tempo, la polarità degli elementi che compongono la frase.



Here are two examples that both express the idea, "The dog eats a meaty bone." The value labeled 'b1' in each example is a compound value.

```
(e1 / eat
:subject (d1 / dog)
:object (b1 / bone
:premod (m1 / meaty)))

(e1 / eat
:agent (d1 / dog)
:patient (b1 / bone
:premod (m1 / meaty)))
```

"I see a man with a telescope".

Due interpretazioni semantiche in Halogen

```
(*OR* (a1 / say
:agent (j1 / "John")
:saying (*OR*
(s1 / (*OR* see watch)
:agent I
:patient (m1 / man
:accompanier (t1 / telescope)))
(s2 / see
:agent I
:patient (m2 / man)
:instrument (t1 / telescope)))) (a2 /
sing
:agent (j2 / "Jane")
:saying (*OR* s1 s2)))
```

John said it

or

Jane sang it

L'uomo
portava il
telescopio

or

Il telescopio
come
strumento
Per guardare

35

Software per la generazione di speech (TTS)

Festival (free):

Center for SpeechTechnology Research, University of Edinburgh.
<http://www.cstr.ed.ac.uk/projects/festival/download.html>

Mary (free)

<http://mary.dfki.de/>

Emofilt

<http://emofilt.syntheticspeech.de/>

Loquendo TTS: Telecom Italia

22 lingue, voci maschili e femminili, con intonazioni emotive; anche in versione embedded

<http://www.loquendo.com>

Una demo qui: http://www.loquendo.com/en/demos/demo_tts.htm

MIT-Spoken Language Systems

<http://groups.csail.mit.edu/sls/about/>

Un TTS in italiano, al CNR di Padova

<http://www2.pd.istc.cnr.it/TTS/It-TTS.htm>

Una demo qui: <http://www2.pd.istc.cnr.it/FESTIVAL/home/demo-interactive.htm>

36

Mary: un generatore di speech 'emotivi'

The "reference" publication for MARY is this one:

M. Schröder & J. Trouvain (2003). The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. International Journal of Speech Technology.

It is available via

<http://mary.dfki.de/documentation/publications>

and via my publications page,

<http://www.dfki.de/~schroed/publications.html>

The best way to understand how MARY interprets its input may be to look at how, e.g., APML is converted into RAWMARYXML.

The best for students would be to ask questions on the mary-users mailing list

<http://www.dfki.de/mailman/listinfo/mary-users>

37

Software per il NLU

Numerosi software per il parsing.

Ad esempio: Phoenix (che è parte del simulatore di dialogo Galaxy)

Sistemi di ASR (automatic speech recognition)

Loquendo

<http://www.loquendo.com/en/technology/asr.htm>

Mediavoice

<http://www.mediavoice.it/index.asp>

38

Software semplici per lo sviluppo di dialoghi speech-based

VoiceXML: Voice Extensible Markup Language

Molti TTS utilizzano VoiceXML come standard

Un Tutorial qui: <http://www.voicexml.org/tutorials/intro1.html>

VoiceXML (VXML) is the W3C's standard XML format for specifying interactive voice dialogues between a human and a computer.

It allows voice applications to be developed and deployed in an analogous way to HTML for visual applications.

Just as HTML documents are interpreted by a visual web browser,

VoiceXML documents are interpreted by a voice browser.

VoiceXML has tags that instruct the voice browser to provide speech synthesis, automatic speech recognition, dialog management, and audio playback.

An example of a VoiceXML document:

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

```
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml"> <form>  
<block> <prompt> Hello world! </prompt> </block> </form> </vxml>
```

39

VoiceXML is designed for creating audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of spoken input, telephony, and mixed initiative conversations.

Un esempio

```
<?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">  
<form>  
<field name="drink">  
<prompt> Would you like coffee, tea, milk, or nothing? </prompt>  
<grammar src="drink.grxml" type="application/srgs+xml"/>  
</field>  
<block>  
<submit next="http://www.drink.example.com/drink2.asp"/>  
</block>  
</form>  
</vxml>
```

40

Related standards

SRGS and SISR

The Speech Recognition Grammar Specification (SRGS) is used to tell the speech recognizer what sentence patterns it should expect to hear: these patterns are called grammars. Once the speech recognizer determines the most likely sentence it heard, it needs to extract the semantic meaning from that sentence and return it to the VoiceXML interpreter. This semantic interpretation is specified via the Semantic Interpretation for Speech Recognition (SISR) standard. SISR is used inside SRGS to specify the semantic results associated with the grammars, i.e., the set of ECMAScript assignments that create the semantic structure returned by the speech recognizer.

SSML

The Speech Synthesis Markup Language (SSML) is used to decorate textual prompts with information on how best to render them in synthetic speech, for example which speech synthesizer voice to use, when to speak louder or softer.

PLS

The Pronunciation Lexicon Specification (PLS) is used to define how words are pronounced. The generated pronunciation information is meant to be used by both speech recognizers and speech synthesizers in voice browsing applications.

CCXML

The Call Control eXtensible Markup Language (CCXML) is a complementary W3C standard. A CCXML interpreter is used on some VoiceXML platforms to handle the initial call setup between the caller and the voice browser, and to provide telephony services like call transfer and disconnect to the voice browser. CCXML can also be used in non-VoiceXML contexts such as teleconferencing.

RAD: Rapid Application Developer

Centre For Spoken Language Understanding, University of Oregon

A software package which enables very fast development of *simple spoken dialogue systems*.

The system comes provided with the text-to-speech system Festival and its own speech recognition system.

The latter is a speaker-independent speech recognition system, which means that it is not possible to train it for an individual speaker. In general, one expects a speaker-independent speech recogniser to have worse performance than a system that can be trained for a particular person, so RAD applications generally build in quite strong expectations about what a person will say at any point in the dialogue.

Documentation available at:

<http://www.cslu.ogi.edu/toolkit/docs/2.0/apps/rad/index.html>

42

Software più complessi per la simulazione di dialoghi

Galaxy (MITRE Corporation)

<http://communicator.sourceforge.net>

TRINDIKIT (free)

In Prolog. Risultato di un Progetto Europeo.

Diversi sistemi utilizzano Galaxy o Trindikit integrandoli con altre componenti (ad esempio, per la realizzazione di dialoghi speech-based)

43

CMU Communicator

Carnegie Mellon University

<http://www.speech.cs.cmu.edu/Communicator/index.html>

The system allows to construct multi-destination travel itineraries, consisting of air travel, hotel, and car reservations.

It knows about most North American and some European routes. It currently provides up-to-date flight information, and you can ask it for the cost of an itinerary.

The Travel Planning systems uses the *Galaxy architecture* for inter-module communication and for logging.

The core modules of the system are implemented in C/C++ and use native Galaxy messages to communicate information.

Some of the modules, in particular ABE, are implemented in Perl. We found it convenient to isolate these modules, so the actual implementation consists of the module itself plus a Galaxy-based proxy that manages communication between the module and the rest of the (Galaxy-based) system.

44

CMU-Interpretazione delle mosse dell'utente

Sphinx decodes the input utterance to produce a top-1 hypothesis and adds a confidence marker to each word in the hypothesis. Sphinx interacts with other modules. Specifically, it will, on barge-in, send a signal to the synthesis module to stop speech output; it will accept language-model switching messages from the dialog manager and will send the final decoding to Phoenix.

Phoenix parses the decoding it receives from Sphinx, using a semantic grammar. Phoenix can potentially produce multiple parses. Parses consist of hierarchical slot arrays, with the slots corresponding to semantic entities in the domain ontology. The ontology is simple and is best thought of as a type hierarchy. The structure of the grammar directly mirrors the ontology and additionally includes (domain-independent) discourse concepts, for example "yes" and "no". Together these define the expected user language for the domain.

Helios is the "post-parser". Its role is to assess the level of confidence for an incoming parse using information from the decoder, parse and dialog levels of the system. The accordingly annotated parse is then sent to the dialog manager.

45

CMU-Dialog Manager

The Communicator Dialog Manager implements the AGENDA dialog manager.

The module contains an *execution engine*, and a *handler library*.

The library is domain-specific and contains individual handlers and handler (sub-)trees, both of which are assembled into a dynamic *product* tree over the course of a session.

The Engine additionally manages the dialog *agenda* which controls the interpretation of user inputs.

Handlers are implemented as C++ objects and incorporate logic for interpreting particular inputs, interacting with domain agents or for managing child-nodes in the product tree.

The product is built up dynamically over the course of a session; as a consequence the system does not follow a dialog "script" in the conventional sense, rather the sequence of interactions is determined by (legal) extensions to the product and by user topic-focusing behavior. The Dialog Manager focuses on the *task* and *discourse* aspects of the dialog and performs minimal domain-specific reasoning, which is primarily located in the ABE module.

46

CMU-Domain-specific reasoning

ABE performs a variety of domain-specific functions and is the "application" that the dialog system interfaces to. The functions include *access to information in the system database*, *retrieval of information on the web* and *domain-specific reasoning*.

ABE interfaces to web-based resources to obtain information about flights and hotels (schedules and prices for flights and locations, prices and availability for hotels).

ABE also incorporates domain-specific reasoning to deal with, for example, the resolution of ambiguous references ("Is that Portland in Maine or Portland in Oregon?") and managing solution sets (for example, ranking flights on "desirability").

ABE interacts with the database, which contains geographical information (about 500 world-wide destinations) and information about airlines. The database also contains information about how users might refer to various entities in the domain (for example airport names) and information about how the system should in turn refer to entities when speaking to the user.

47

MIT-Spoken Language Systems

To support its research on spoken language systems for human/computer interaction, the SLS group has developed its own suite of core speech technologies. These technologies include:

- speech recognition (SUMMIT)
- natural language understanding (TINA)
- dialogue modeling
- language generation (GENESIS)
- speech synthesis (ENVOICE)

The core speech technologies can be integrated to create conversational systems using the GALAXY architecture for conversational speech systems. The creation of GALAXY has enabled developers to rapidly create conversational systems for a wide variety of applications.

48

Dialogue Corpora

Nuovi metodi possono essere testati su corpora già raccolti.
Elencati, ad esempio, in:

The Dialogue Diversity Corpus
(W.C Mann, 2003)

<http://www-rcf.usc.edu/~billmann/diversity/DDivers-site.htm>

Domini:

Tutoring
Transport planning
Interaction between friends
Computer help

49

Introduzione

Unità 1: Ragionamento logico:

- Formalizzazione
- Risoluzione

Unità 2: Generazione di linguaggio naturale

- Teorie
- Metodi

Unità 3: Ragionamento incerto

- Reti Causali Probabilistiche
- Reti dinamiche

Unità 4: Modelli di Utente

- Modelli logici
- Modelli con incertezza

Unità 5: Comprensione del linguaggio naturale

Unità 6: Simulazione di dialoghi

- Modelli basati su ATN
- Modelli basati su IS

Unità 7: Affective Computing

Programma del Corso

50

Collegamenti con altri Corsi

Prerequisiti: i contenuti dei Corsi di:

Interazione Uomo-Macchina 1 (laurea triennale)

Nozioni di base di logica

Nozioni di base di teoria della probabilità

Nozioni di base sulle Grammatiche

Approfondimenti utili:

Informatica Grafica

Elaborazione di segnali

Sistemi ad Agenti

Reti Neurali

Possibili sovrapposizioni parziali (che cercheremo di evitare)

Ingegneria della Conoscenza

51

Materiale Didattico

- I lucidi verranno messi in linea, possibilmente prima di ogni lezione: lo studio dei lucidi è **necessario** ma *non sufficiente* ai fini dell'esame.
- Verrà organizzata una pagina web che contiene:
 - un elenco di **articoli** di cui si suggerisce la lettura
 - il materiale relativo alle **esercitazioni di laboratorio**
- Altri approfondimenti possibili verranno suggeriti durante il Corso, con link a siti rilevanti.

52

Modalità di Esame

- Una **prova scritta**, individuale, sugli argomenti trattati a lezione (teoria ed esercizi).
- Una **prova orale** (di gruppo) nella quale vengono discussi:
 - gli articoli studiati
 - gli argomenti eventualmente approfonditi mediante accesso ai siti web suggeriti
 - Le Esercitazioni di Laboratorio sviluppate

53

Primi link sulle interfacce intelligenti

Il punto di vista di Henry Lieberman (MIT, Boston) sulle interfacce intelligenti:

<http://web.media.mit.edu/~lieber/Teaching/Int-Int/Int-Int-Intro.html>

Una lista dei siti web sull'argomento e una miniera di link a materiale sull'argomento:

<http://www.aaai.org/aitopics/pmwiki/pmwiki.php/AITopics/Interfaces#web>

54