

NATURAL LANGUAGE PROCESSING

Introduzione

Il linguaggio naturale è il linguaggio parlato dagli umani (Italiano, Inglese, ...).

Esso viene chiamato naturale al fine di distinguerlo dai linguaggi artificiali. Un esempio di linguaggio artificiale è rappresentato dai linguaggi di programmazione.

Il linguaggio naturale viene utilizzato in modo naturale, cioè senza rendersi conto che effettivamente vengono seguite regole linguistiche.

L'Elaborazione del Linguaggio Naturale (Natural Language Processing, NLP) è una disciplina che ha come scopo quello di fornire metodologie per programmare le macchine al fine di elaborare grandi quantità di dati rappresentati mediante linguaggio naturale (esempi: conversazione orale, testo scritto). Tali metodologie possono poi essere utilizzate per diverse finalità, come ad esempio quella di far eseguire alcune attività, prima eseguite dagli umani, alle macchine.

Affinché il NLP abbia successo, è fondamentale il Natural Language Understanding (NLU). Il NLU ha come scopo quello di far comprendere (ad esempio alle macchine) il significato e il contenuto di una conversazione o di un testo.

NATURAL LANGUAGE PROCESSING

Introduzione

Alcune applicazioni del NLP sono:

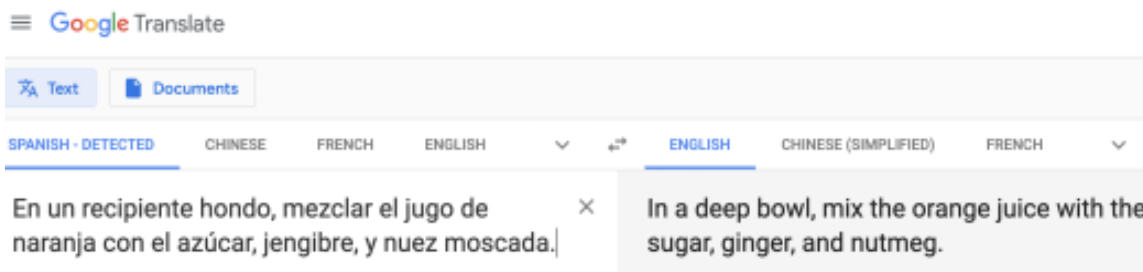
- Machine Translation
- Information Retrieval
- Question Answering
- Dialogue Systems
- Speech Recognition (ASR e TTS)

NATURAL LANGUAGE PROCESSING

Machine Translation

Per Machine Translation (MT) si intende l'utilizzo dei computers per effettuare traduzioni da un linguaggio a un altro linguaggio.

Il MT può essere riferito a differenti applicazioni pratiche. L'uso attuale più comune del MT riguarda l'accesso alle informazioni (*information access*): si pensi ad esempio alla traduzione di alcune frasi sul web.



I miglioramenti nel MT possono aiutare a ridurre il digital divide per quanto riguarda l'accesso alle informazioni; ad esempio, le ricerche in lingua Inglese restituiscono più informazioni rispetto a ricerche svolte in altre lingue.

Un altro utilizzo comune del MT è il supporto agli esseri umani che svolgono attività di traduzione (Computer-Aided Translation, CAT). Ad esempio i sistemi MT possono fornire una bozza che poi viene opportunamente editata da un umano.

NATURAL LANGUAGE PROCESSING

Information Retrieval

L'Information Retrieval (IR) si occupa della rappresentazione, organizzazione e memorizzazione di informazioni in base ai bisogni dell'utente.

Un'applicazione dell'IR è l'*ad hoc retrieval*, nel quale un utente fa una *query* ad un sistema IR che restituisce un insieme ordinato di documenti contenuti in collezioni.

Per documento si intende un'unità di testo come ad esempio una pagina web o un articolo scientifico.

Una collezione si riferisce a una serie di documenti utilizzati per soddisfare le richieste dell'utente.

Una *query* rappresenta il bisogno di informazione dell'utente espresso come un insieme di termini.

Un termine si riferisce ad una parola in una collezione.

NATURAL LANGUAGE PROCESSING

Question Answering

Il Question Answering (QA) è un'applicazione che ha lo scopo di rispondere a domande formulate da utenti.

Question	Answer
Where is the Louvre Museum located?	in Paris, France
What are the names of Odin's ravens?	Huginn and Muninn
What kind of nuts are used in marzipan?	almonds
What instrument did Max Roach play?	drums
What's the official language of Algeria?	Arabic

Un ambito del QA è il retrieval-based QA, nel quale, per rispondere a una domanda di un utente, prima si cercano brevi segmenti di testo nel web o in altre collezioni di documenti; i risultati di tale ricerca vengono poi utilizzati per generare una risposta.

NATURAL LANGUAGE PROCESSING

Dialogue Systems

La conversazione o dialogo è l'ambito più importante del linguaggio. È il primo tipo di linguaggio che impariamo da bambini e inoltre lo utilizziamo costantemente.

Si possono distinguere due tipi di architetture:

- *Task-oriented dialogue systems*: tali sistemi conversano con gli utenti per l'esecuzione di compiti predefiniti come ad esempio il controllo di applicazioni o la ricerca di ristoranti; essi si basano su una struttura dati chiamata *frame*, la quale rappresenta la conoscenza che un sistema deve acquisire dall'utente (per esempio l'orario al quale restituire un segnale di allarme per la sveglia).
- *Chatbots*: essi sono progettati per riprodurre conversazioni lunghe e non strutturate (o «chats») caratteristiche dell'interazione umano-umano.

Sistemi moderni includono aspetti di entrambe le architetture: *chatbots* industriali, come ad esempio ChatGPT, possono condurre lunghe conversazioni non strutturate; *digital assistants* come Siri o Alexa sono generalmente *frame-based dialogue systems*.

COMUNICAZIONE VERBALE

Speech Dialogs – Introduzione

Le interfacce basate sul parlato sono molto diffuse perché:

- Il parlato è la forma più naturale della comunicazione tra gli umani.
- Tali interfacce richiedono spazio limitato, ad esempio un microfono e un altoparlante possono essere posizionati anche in dispositivi indossabili. Tuttavia, ciò richiede uno sforzo computazionale significativo.
- Le mani e gli occhi vengono lasciati liberi: tale aspetto permette un utilizzo molto diffuso dell'interazione basata sul parlato in molte applicazioni che richiedono il controllo di un dispositivo da parte dell'utente (esempio: guida della macchina).

COMUNICAZIONE VERBALE

Speech Dialogs – Introduzione

Utilizzando il parlato come modalità di input e output si ottengono i *dialog systems* basati sul parlato. In generale questi sistemi sono, in ordine gerarchico di complessità:

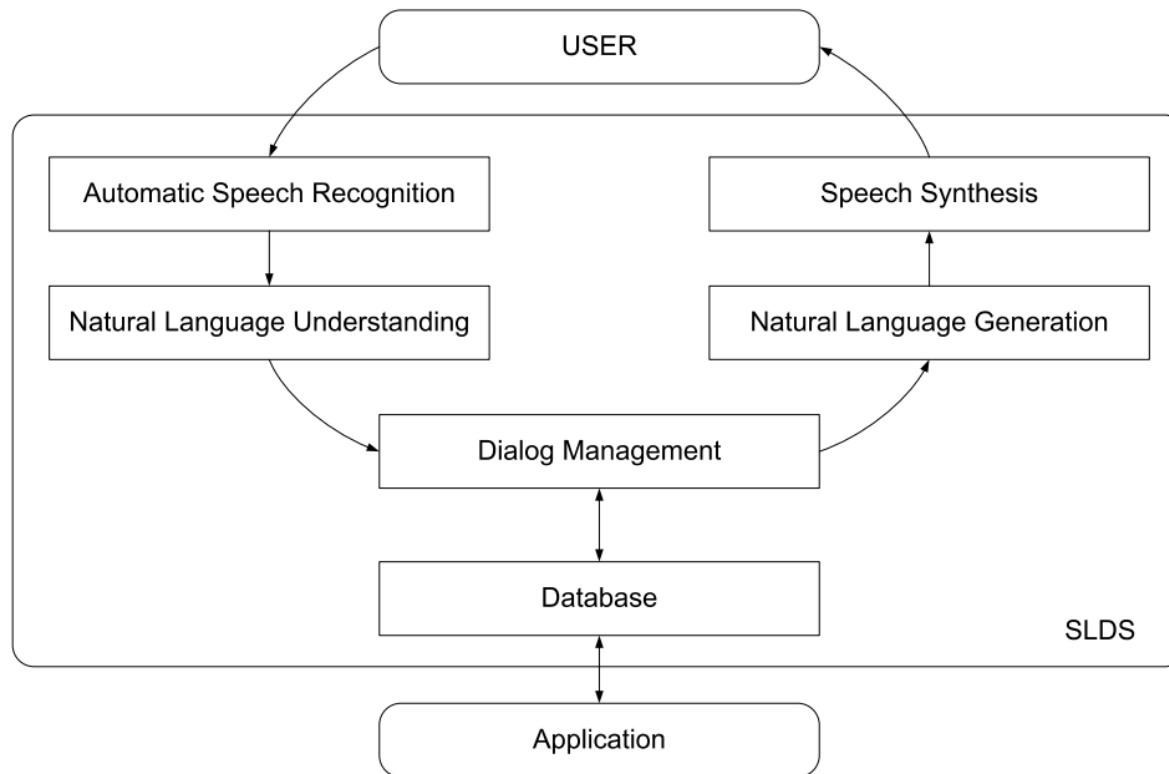
- sistemi che permettono il controllo (esempio: guida della macchina);
- sistemi che permettono l'IR (esempio: richiesta di informazioni su un volo);
- sistemi che permettono l'esecuzione di transazioni strutturate (*structured transactions*) con l'utilizzo della voce (esempio: controllo delle merci);
- sistemi che permettono l'esecuzione di task caratterizzati dalla combinazione di IR e transazioni (esempio: prenotazione di un hotel tenendo conto delle disponibilità dei voli);
- sistemi che permettono l'esecuzione di task complessi (*complex tasks*), come ad esempio l'assistenza alle persone anziane.

Più specificatamente, un dialogo può essere definito come uno scambio di differenti elementi in una conversazione reciproca tra almeno due istanze che possono essere umani o macchine con almeno un cambio dello speaker.

COMUNICAZIONE VERBALE

Speech Dialogs – Introduzione

Il dialogo basato sul parlato viene indicato come *speech dialog*. Più specificatamente si parla di *Spoken Language Dialog Systems (SLDS)*, i quali generalmente sono una combinazione di riconoscimento del parlato (ASR), Natural Language Understanding (NLU), generazione del dialogo e sintesi del parlato.



Esempio di architettura di un SLDS.

COMUNICAZIONE VERBALE

Speech Dialogs – Introduzione

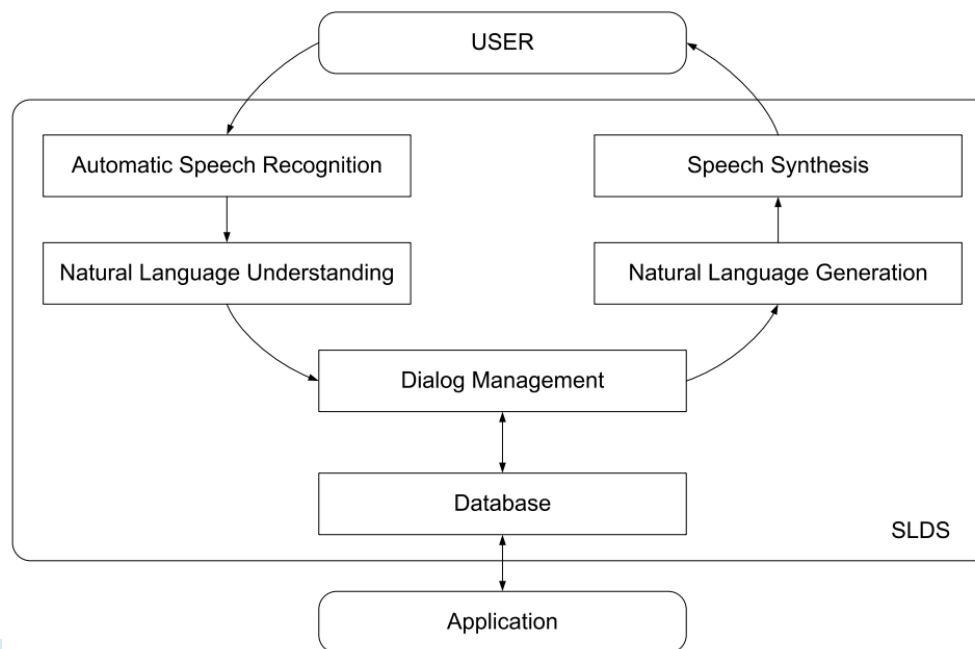
L'utente (parte superiore della figura della slide precedente) controlla un'applicazione (parte inferiore della figura della slide precedente); il *SLDS* agisce da interfaccia. I singoli componenti del *SLDS* sono:

- *Sistema ASR* (si vedano le slide precedenti): tale sistema può essere esteso tenendo conto anche delle espressioni dell'utente in modo da includere aspetti legati alle emozioni.
- *NLU* (si vedano le slide precedenti): interpretazione del significato o dello scopo del contenuto associato al parlato.
- *Dialog Management (DM)*: il *DM* è il modulo principale di un'interfaccia vocale poiché esso svolge la funzione di agente intermedio tra l'utente e l'applicazione ed è responsabile dell'interazione tra loro. Esso opera sulla base di una rappresentazione del significato o dello scopo fornita dal *NLU* che modella cosa l'utente stia presumibilmente dicendo. Sulla base di questa informazione, il *DM* ha diverse opzioni, come ad esempio il cambio dello stato dell'applicazione in caso di controllo vocale o la ricerca di dati in un database di interesse nel caso di un servizio per la ricerca di informazioni. Inoltre, il *DM* decide il tipo di output vocale e quando tale output deve essere prodotto.

COMUNICAZIONE VERBALE

Speech Dialogs – Introduzione

- *Database (DB)*: esso memorizza le informazioni riguardanti il contenuto del dialogo.
- *Natural Language Generation (NLG)*: formulazione della risposta astratta fornita dal *DM*. Esistono differenti approcci, come ad esempio approcci probabilistici e approcci basati su stringhe predefinite.
- *Speech Synthesis (TTS)*: produzione vocale basata sulla risposta formulata dal sistema.



COMUNICAZIONE VERBALE

Speech Dialogs – Strategie di iniziativa

Può essere eseguita una classificazione basata sulla strategia di iniziativa nel dialogo:

- *system-driven*: il sistema ha l'iniziativa nel corso dell'intero dialogo;
- *user-driven*: l'utente ha l'iniziativa nel corso dell'intero dialogo;
- *mixed initiative*: l'iniziativa cambia durante il dialogo.

Di solito la strategia di iniziativa è determinata dal tipo di applicazione e dal tipo di interfaccia vocale. Per esempio, sistemi con un vocabolario limitato tendono ad utilizzare strategie *system-driven*. Il sistema quindi pone domande molto specifiche e l'utente può solo rispondere ad esse. Si deve adoperare tale strategia nel caso di un vocabolario limitato poiché l'insieme degli input è molto limitato. Invece applicazioni *C&C (Command and Control)* adottano strategie *user-driven*: il sistema deve attendere l'input da parte dell'utente prima di poter fare qualcosa.

COMUNICAZIONE VERBALE

Speech Dialogs – Strategie di iniziativa

Tuttavia, sarebbe preferibile poter utilizzare strategie *mixed initiative* nelle quali entrambe le parti possono avere l'iniziativa quando opportuno in base allo stato corrente del dialogo. Quindi l'utente può dialogare con il sistema come farebbe con un altro essere umano.

Il progetto di strategie *mixed initiative* è più difficile per almeno due motivi. Esso è tecnicamente impegnativo, poiché l'utente dovrebbe avere la libertà di poter dire qualunque cosa in qualunque momento: ciò rappresenta una complicazione significativa dal punto di vista del riconoscimento del parlato e dal punto di vista dell'elaborazione del linguaggio. Inoltre, esso è complesso dal punto di vista del dialogo, poiché l'iniziativa deve essere monitorata e le reazioni devono essere flessibili.

COMUNICAZIONE VERBALE

Speech Dialogs – Modelli

All'interno del blocco *Dialog Management* (si vedano le slide precedenti), la struttura della comunicazione viene modellizzata mediante un modello del dialogo. Le due categorie principali di modelli sono:

- *structural models*: viene utilizzato un percorso più o meno predefinito.
- *nonstructural models*: si basano sui principi generali del dialogo.

Finite State Model

La struttura del dialogo rappresenta una sequenza finita di passi o stati predefiniti. Un esempio di *finite state model* è l'automa a stati finiti deterministico. Esso è un grafo nel quale i nodi rappresentano le domande del sistema, le uscite del sistema e le azioni del sistema mentre gli archi mostrano tutti i possibili percorsi che sono etichettati con le espressioni dell'utente. Poiché in un automa a stati finiti deterministico tutti i possibili percorsi sono fissati, esso non permette un controllo esplicito del dialogo. Il dialogo quindi tende ad essere rigido e tutte le reazioni agli input dell'utente sono stabilite. Tale modello può essere formulato come una quintupla

$$\{S, s_0, s_f, A, \tau\}$$

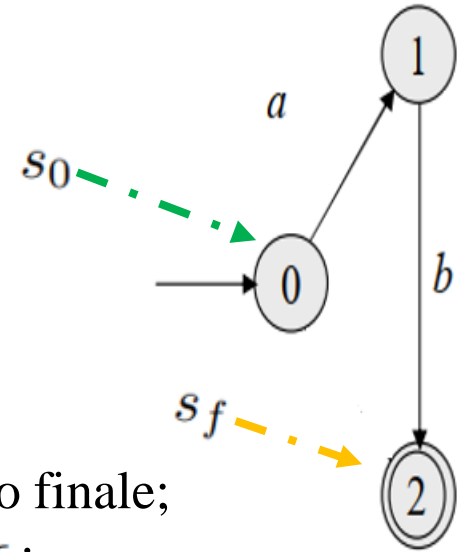
COMUNICAZIONE VERBALE

Speech Dialogs – Modelli

$$\{S, s_0, s_f, A, \tau\}$$

dove:

- s_0 è lo stato iniziale;
- s_f è lo stato finale;
- S è l'insieme degli stati, compresi lo stato iniziale e lo stato finale;
- A è l'insieme delle azioni (eventi), inclusa l'azione vuota ϵ ;
- τ è una funzione di transizione con $\tau : S \times A \rightarrow S$. Con τ si specifica a quale stato conduce un'azione dato lo stato attuale.



Un dialogo è definito come il percorso da s_0 a s_f all'interno dello spazio degli stati. Generalmente, gli automi a stati finiti deterministici rappresentano il modo più naturale per rappresentare dialoghi basati su strategia di iniziativa *system-driven*. Inoltre, sono adatti per uno scambio di informazioni completamente prevedibile grazie al fatto che l'intero modello di dialogo viene definito al momento dello sviluppo. Spesso è possibile rappresentare il flusso del dialogo graficamente, il che implica un processo di sviluppo molto naturale.

COMUNICAZIONE VERBALE

Speech Dialogs – Modelli

Tuttavia non c'è flessibilità, poiché tali modelli sono completamente guidati dal sistema e non sono ammesse deviazioni di percorso. L'input dell'utente è ristretto a singole parole o frasi che forniscono risposte a richieste del sistema accuratamente progettate.

Le attività che richiedono negoziazioni non possono essere modellizzate mediante automi a stati finiti deterministici a causa dell'incertezza del loro risultato. In conclusione, i modelli basati su automi a stati finiti deterministici sono appropriati per attività semplici con menù non profondi e liste corte di opzioni.

Slot filling

Nei sistemi *frame-based*, all'utente possono essere poste domande che abilitano il riempimento di slot (*slot filling*) in un template con lo scopo di eseguire un'attività o un compito. Questi modelli possono essere utilizzati in sistemi IR basati su database come ad esempio la ricerca di informazioni riguardanti il cinema o i voli. Rispetto al modello basato sugli automi a stati finiti deterministici, il flusso del dialogo non è prestabilito ma dipende dal contenuto dell'input fornito dall'utente e dalle informazioni che il sistema deve ricavare.

COMUNICAZIONE VERBALE

Speech Dialogs – Modelli

Se l'utente fornisce più input di quelli richiesti in una certa fase, il sistema è in grado di accettare questa informazione e di verificare se sono necessari altri elementi. Un componente necessario è quindi il *frame*, il quale controlla i segmenti di informazione già formulati.

Se più slot devono ancora essere riempiti, sorge la questione di come trovare l'ordine ottimo rispetto alla richiesta del sistema. Una soluzione può essere quella di limitare l'insieme originariamente ampio di opzioni a priori associate alle azioni in modo da accelerare il dialogo.

Slot	Type	Example Question
ORIGIN CITY	city	"From what city are you leaving?"
DESTINATION CITY	city	"Where are you going?"
DEPARTURE TIME	time	"When would you like to leave?"
DEPARTURE DATE	date	"What day would you like to leave?"
ARRIVAL TIME	time	"When do you want to arrive?"
ARRIVAL DATE	date	"What day would you like to arrive?"

Esempio di frame in un sistema di dialogo frame-based. Sono stati riportati il tipo di ogni slot e una domanda di esempio per il riempimento di ogni slot.

COMUNICAZIONE VERBALE

Speech Dialogs – Modelli

Un vantaggio del controllo del dialogo tramite l'utilizzo di *frame* è l'alta flessibilità per l'utente e la possibilità di riempire slot multipli. I dialoghi sono più corti ed è possibile utilizzare una strategia di iniziativa *mixed initiative*.

Tuttavia, è richiesta una grammatica estesa per poter avere una maggiore flessibilità nelle espressioni degli utenti; inoltre l'algoritmo di controllo del dialogo deve specificare le azioni successive del sistema sulla base dei *frame* disponibili.

Stochastic Model

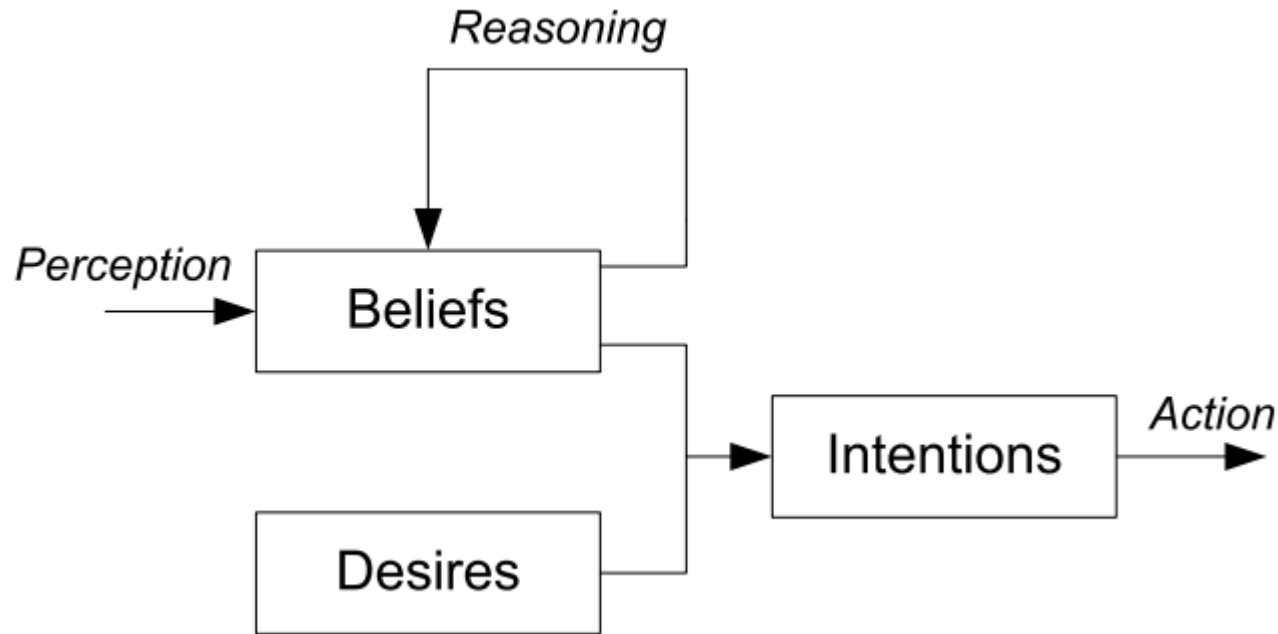
Essi mirano a superare problematiche come la scarsa portabilità a nuovi domini e la limitata predicibilità di potenziali problemi all'interno del processo di progettazione.

Goal Directed Processing

I modelli precedentemente descritti sono adatti per applicazioni *C&C* (*Command and Control*) e per IR, ma sono meno adatti per applicazioni dove l'utente e il sistema devono cooperare per risolvere un problema. In tali applicazioni si utilizza ad esempio il modello *Beliefs, Desires, and Intentions* (*BDI*), riportato nella figura della slide seguente.

COMUNICAZIONE VERBALE

Speech Dialogs – Modelli



Modello BDI.

In tale modello, gli agenti hanno delle opinioni (*beliefs*) sullo stato attuale del mondo e dei desideri (*desires*) su come vorrebbero che fosse. Sulla base di questi elementi, essi determinano le loro intenzioni (*intentions*), cioè il loro obiettivo, e costruiscono un piano composto da azioni per soddisfare i loro desideri.

Anche tali sistemi hanno degli svantaggi, come ad esempio l'alto costo computazionale.

COMUNICAZIONE VERBALE

Speech Dialogs – Principi di progettazione

Nel progetto di sistemi di dialogo basati sul parlato, solitamente vengono seguiti i principi dell'*user-centered design*:

- Studio dell'utente e del task: comprensione della tipologia di utenti e del task mediante interviste ai potenziali utenti, studio di sistemi con caratteristiche simili e studio dei dialoghi umano-umano associati.
- Costruzione di simulazioni e prototipi: nella costruzione di sistemi di dialogo, uno strumento cruciale è il *Wizard-of-Oz system*. Mediante tale strumento, l'utente interagisce con un'interfaccia software in modo da poter testare una determinata architettura prima dell'implementazione.
- Procedura iterativa di test sugli utenti: il test sugli utenti è essenziale nella progettazione. Inoltre è importante considerare il *value sensitive design*, nel quale vengono considerati, durante la progettazione, i possibili benefici, danni e stakeholders del sistema realizzato.

Riferimenti Bibliografici

- [1] Kraiss, K. -F. (2006). Advanced Man-Machine Interaction: Fundamentals and Implementation. Springer-Verlag Berlin Heidelberg. ISBN-10: 3-540-30618-8
- [2] Daniel Jurafsky, James H. Martin (2024). Speech and Language Processing – An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.