

# Analisi e specifiche per il supporto dello Screen Reader Semantico

## Survey

Nel presente documento si riassumono i contenuti del deliverable **"D4.1\_1 Analisi e specifiche per il supporto dello Screen Reader Semantico"** relativo all'attività *RI 4.1 Analisi sull'estensione per il supporto dello Screen Reader Semantico*, nell'ambito del quarto Obiettivo Realizzativo (OR 4) "Studio di interfacce multimodali avanzate e dispositivi speciali per i disabili".

L'obiettivo di SAPI è quello di dare anche ad un'utenza con disabilità visiva la possibilità di usufruire dei servizi messi a disposizione da Poste Italiane. Per tale motivo l'attività di ricerca è stata orientata ad individuare le soluzioni assistive più consone e soprattutto più efficaci per tali utenti. La presenza di una patologia visiva limita notevolmente l'interazione visuale con la piattaforma e addirittura nel caso peggiore (ovvero quello di cecità assoluta) essa è completamente compromessa. La soluzione proposta da SAPI per chi è affetto da patologie visive (in particolar modo cecità assoluta) e vuole consultare documenti digitali è un Lettore di Schermo "Screen Reader" che consente la navigazione semantica dei documenti attraverso dialoghi vocali. Lo Screen Reader SAPI è un qualcosa di veramente innovativo nel campo delle tecnologie assistive in quanto coniuga la potenza di un'applicazione multimodale con le funzionalità di un lettore di schermo tradizionale (degnamente rappresentato da JAWS). Ad onor di cronaca, una soluzione di questo genere (denominata Active Reader) è stata già presentata da un'azienda italiana che opera nel settore delle tecnologie vocali. Alcune delle funzionalità pensate per SAPI sono già presenti in Active Reader, altre invece come si vedrà più avanti continuano a conservare il loro carattere distintivo di originalità e di innovazione. Active Reader consente all'utente la fruizione di testi in maniera del tutto simile ad uno Screen Reader ma con in più la possibilità di impartire comandi vocali, in linguaggio naturale, per controllare l'interfaccia dell'applicazione e, soprattutto, la navigazione all'interno del testo. Active Reader elimina in questo modo la dipendenza dalla tastiera, da sempre considerato come fattore "limitante" per i disabili visivi. D'altro canto non implementa tutte le funzionalità tipiche di uno screen reader come JAWS. Non è, in effetti, un vero "lettore di schermo". La sua funzione è, per ora, solo quella di leggere documenti. Il software legge puro testo (TXT) ma è anche in grado di leggere documenti creati con la suite Microsoft Office 2007. Le principali funzionalità di Active Reader sono le seguenti:

- **Meta informazioni:**

- Lettura automatica iniziale delle informazioni sul documento: numero pagine,... ecc.
- Indicazione della posizione dell'utente all'interno del documento: l'utente chiede "Dove sono?", l'applicazione risponde, per esempio: "Sei al paragrafo 5 di 12 paragrafi della pagina 15 di 250".

- **Navigazione vocale del documento:**
  - Salti per (n) pagine, (n) paragrafi, (n) righe e (n) parole. Per esempio: "Rileggi la riga precedente", "Salta al paragrafo successivo",... ecc.
  - Inserimento, rimozione e salto tra segnalibri all'interno del documento.
  - Emulazione della scansione del testo: il software legge (n) parole per paragrafo, simulando la modalità di lettura a scansione, impossibile per un non vedente.
  
- **Controllo vocale dell'interfaccia grafica:**
  - Evidenziazione del testo letto in tempo reale (utile per utenti ipovedenti ma anche per utenti dislessici).
  - Modifica della dimensione dei caratteri.
  - Impostazione di diverse modalità a contrasto elevato.
  - Navigazione vocale dell'interfaccia dell'applicazione: menu,... ecc.
  
- **Controllo vocale dell'interfaccia vocale:**
  - Gestione del volume del TTS.
  - Gestione della velocità di lettura.

Cosa aggiunge lo Screen Reader SAPI all' *Active Reader* ? Quali vantaggi ulteriori offre agli utenti ipovedenti e non vedenti?

Una prima risposta potrebbe essere la seguente: lo Screen Reader SAPI potrebbe aggiungere alle feature di Active Reader tutte quelle funzionalità tipiche dei *libri parlanti* non implementate in Active Reader. Tuttavia in SAPI non ci si vuol limitare a trasporre nello Screen Reader l'involuppo delle funzionalità già realizzate in applicazioni simili; si vuole invece dare in termini prestazionali un contributo innovativo che rappresenti un significativo passo avanti nell'ambito delle attività di ricerca dedicate ai tool assistivi per gli ipovedenti e i non vedenti: la navigazione semantica (per concetti) e tematica predittiva.

### **Vantaggi della navigazione semantica e tematica predittiva**

Si tratta di una modalità di navigazione più pratica ed efficiente che si avvicina ancor più al modo di operare quotidiano dell'utente, vale a dire che:

- è il più possibile efficiente rispetto agli interessi e alle preferenze dell'utente dichiarate e dedotte (information filtering);
- consente all'utente la consultazione veloce (lettura " a volo di uccello") di un documento digitale, in particolare di un eBook, per verificare se è di proprio interesse;
- consenta il browsing tematico e per concetti;
- tenga conto della storia delle interazioni dell'utente in modo da poterne predire il comportamento (User Plan Recognition).

Il sistema deve quindi soddisfare le seguenti esigenze che sono emerse in una prima analisi generale del problema:

- la presentazione delle informazioni contenute in un documento deve avvenire esprimendo adeguatamente agli utenti la struttura logica e semantica del contenuto;
- i contenuti di interesse dell'utente devono essere facilmente individuabili e raggiungibili. Tutti i contenuti devono essere raggiungibili sulla base del loro significato e la loro individuazione deve risultare il più possibile semplice, efficiente ed immediata, riducendo così lo sforzo operativo continuo dell'utente;
- il reperimento delle informazioni desiderate deve essere il più possibile accurato in modo da ridurre la quantità di informazioni superflue, indesiderate e non pertinenti. Il fatto stesso che non viene letto tutto il contenuto del documento, ma solamente quelle parti che contengono l'informazione richiesta, riduce drasticamente la quantità di testo letto all'utente.

Con una navigazione di tipo classico (GUI) l'utente con disabilità visiva non può effettuare le tipiche operazioni di *browsing* così come letteralmente è inteso nell'accezione comune, ossia non può effettuare un'analisi se pur sommaria dei contenuti di un documento per comprendere in pochi istanti quali sono i suoi contenuti principali. Questa operazione di interpretazione e di analisi dovrebbe quindi essere svolta dallo Screen Reader, in modo da fornire all'utente una descrizione generale del documento digitale, presentandogli gli argomenti principali in esso contenuti. In tal modo un utente ipovedente o non vedente è in grado di avere una visione panoramica, a volo d'uccello, almeno sui suoi contenuti tematici in certi casi in maniera anche più rapida di quella che potrebbe avere in modo naturale un utente normodotato (Figura 1):

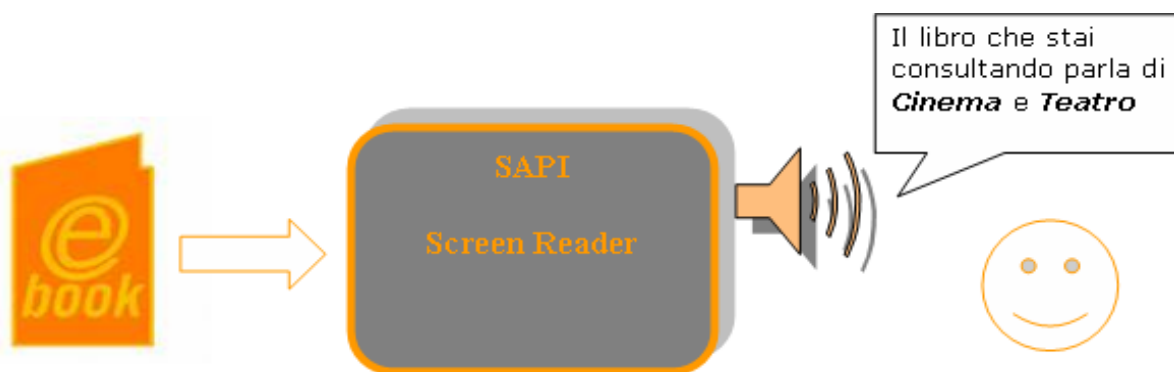


Figura 1: Consultazione veloce dei contenuti tematici di un documento.

In Figura 2 è rappresentato l'approccio architetturale ibrido Client/Server-based adottato per lo Screen Reader SAPI. Durante la fase di installazione dello Screen Reader sul terminale utente, lo stesso potrà essere configurato, in base alle risorse vocali presenti sul Client, per poter funzionare in una delle due modalità. Inoltre, le Speech API saranno realizzate in modo tale da essere compatibili con i principali Speech Engine (Loquendo, Nuance, ...) presenti sul mercato.

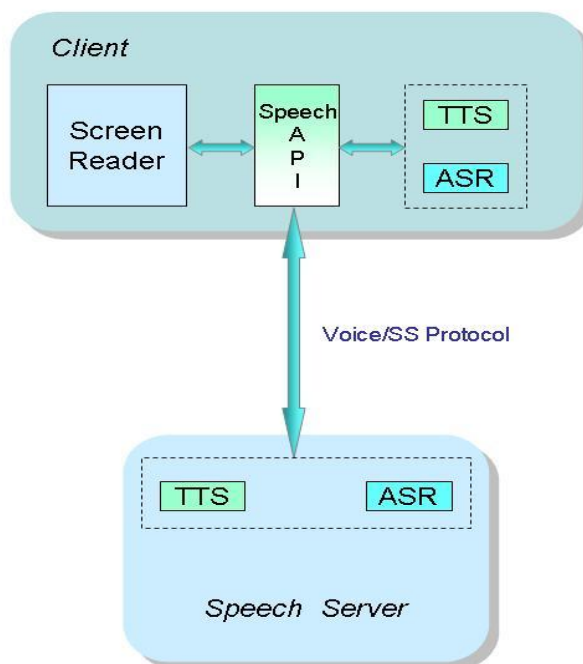


Figura 2: Architettura Ibrida Client/Server-based adottata nello Screen Reader SAPI.

In Figura 3 è rappresentata l'architettura funzionale dello Screen Reader SAPI ed in essa sono riportati i principali moduli funzionali costituenti il tool assistivo. Come si può notare alcuni moduli funzionali (il *Topic Recognizer* e il *Semantic Tagger*) sono gli stessi utilizzati per l'Audio Browser Semantico e perciò valgono le stesse considerazioni ovvero il modulo *Doc Semantic Tagger* (presente nel riquadro tratteggiato) che potrebbe essere utilizzato in modo congiunto o in alternativa al modulo *Topic Recognizer* verrà escluso dalla attuale configurazione considerato che, rispetto ai fini progettuali, il suo valore aggiunto non giustifica l'effort richiesto dalla sua complessità realizzativa. L'Analizzatore Semantico comunque non è stato omesso dalla trattazione in quanto si ritiene che esso, così come è stato concepito in SAPI, rappresenti una soluzione di grande interesse per l'analisi morfologica e semantica dei testi ed in particolare per quella orientata al *topic detection*.

A differenza dell'Audio Browser Semantico, strettamente connesso al sistema SAPI per quanto concerne l'acquisizione di alcune informazioni del profilo d'utente dai moduli Profiler (preferenze esplicite) ed Evolver (predizione del comportamento d'utente), lo Screen Reader è stato concepito come tool assistivo a corredo di SAPI completamente autonomo.

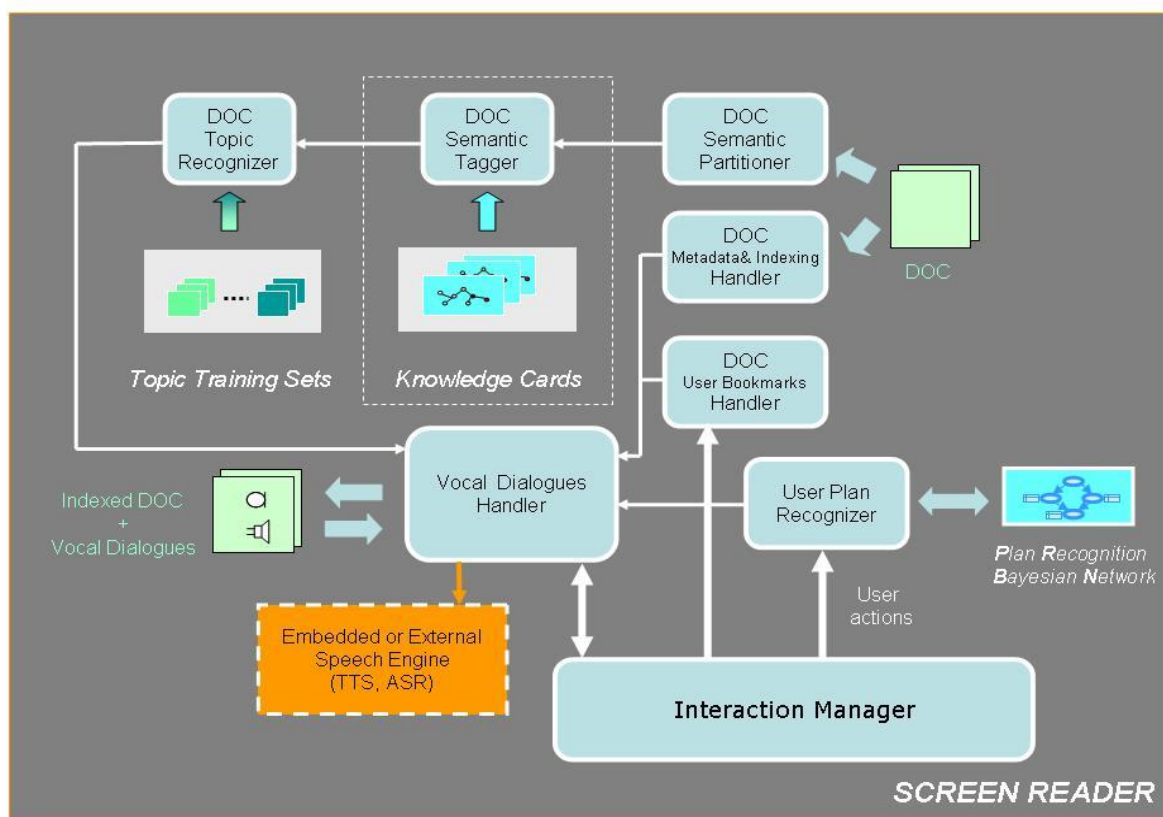


Figura 3: Architettura funzionale dello Screen Reader SAPI.

Le principali funzionalità implementate dallo Screen Reader SAPI sono le seguenti:

- Metadazione ed indicizzazione del documento digitale;
- Analisi semantica (*text segmentation*) del testo;
- Classificazione tematica dei segmenti testuali;
- Generazione automatica dei dialoghi vocali sincronizzati con l'interfaccia GUI (dopo alcune sessioni di consultazione del documento potranno essere disponibili informazioni relative ai segnalibri introdotti dagli utenti e alla predizione della sua prossima richiesta).

✓ **Metadazione ed indicizzazione del documento digitale**

La struttura di un documento elettronico verrà descritta dallo Screen Reader in modo uniforme e semanticamente ben definito attraverso l'utilizzo di un insieme appropriato di metadati. I metadati che verranno utilizzati a tale scopo verranno mutuati in parte da quelli forniti dallo standard Dublin Core, mentre alcuni saranno definiti ad hoc in modo da costituire insieme ai primi un Application profile per SAPI che potrà essere eventualmente utilizzato in futuro anche da altre applicazioni. Verranno presi in considerazione non solo i 15 termini costituenti il Dublin Core Metadata Element Set ma tutti quelli gestiti dalla Dublin Core Metadata Initiative. A valle del processo di annotazione semantica dei documenti elettronici

lo Screen Reader effettuerà l'indicizzazione degli stessi. Per ciascun documento verrà generato un file XML in cui a ciascun elemento (ad. es., Capitolo 2, Riferimenti bibliografici, Introduzione, etc.) del documento richiamabile vocalmente dall'utente sarà associato un indice costituito dalla seguente tripla : < Nr. di pagina, Nr. di riga, Nr. di parola>. Ciò consentirà allo Screen Reader di navigare agevolmente nella struttura della risorsa documentale in base alle richieste dell'utente. Vale la pena notare che ogni elemento indicizzato farà parte della grammatica dinamica che verrà generata dallo Screen Reader per il riconoscimento vocale *speaker independent*.

✓ **Analisi Semantica del testo**

Scopo dell'analisi semantica del testo è quello di partizionare il testo ovvero ciò che in letteratura è noto come *Text Segmentation* o *TextTiling* e vale a dire il processo di elaborazione del testo rivolto ad individuarne i punti di discontinuità semantica, le cosiddette *topic boundaries*. Non rientra negli obiettivi del TextTiling annotare semanticamente, attraverso ad esempio l'assegnazione di un' etichetta tematica (*topic label*), i vari segmenti testuali individuati. Attualmente esistono vari algoritmi di segmentazione del testo; per SAPI si è scelto di utilizzare l'algoritmo di Hearst sia perchè la sua accuratezza soddisfa adeguatamente i requisiti progettuali sia per la sua ampia diffusione e semplicità

✓ **Classificazione tematica dei segmenti testuali**

Scopo della classificazione tematica del testo è la categorizzazione dei testi. La classificazione del testo avviene attraverso l'individuazione delle *features* (caratteristiche del testo in esame), quali ad esempio: la lunghezza del documento, la posizione delle parole, la loro frequenza e il numero di vocaboli. Trovare le dipendenze tra questi parametri, e riuscire a combinarle, grazie a procedimenti indotti dalle probabilità condizionate, porta ad ottenere degli indici confrontabili, per ogni categoria a cui il testo può far riferimento. Le categorie, sono intese come raggruppamenti di documenti appartenenti ad uno specifico argomento. Il classificatore, dato un testo in ingresso, provvede a svolgere i calcoli necessari ad associare il documento in esame ad una delle classi o categorie precedentemente apprese.

In SAPI verrà realizzato un classificatore tematico n-ario per la categorizzazione dei segmenti testuali seguendo un approccio *single-labeled*. Per l'implementazione del classificatore sarà utilizzata la piattaforma di sviluppo open source Mallet che è costituita da librerie sviluppate in JAVA per l'applicazione dei metodi di machine learning al testo. Essa fornisce gli strumenti necessari per un approccio all'elaborazione statistica dei linguaggi naturali, alla classificazione dei documenti, al clustering, all'information extraction, e ad altre applicazioni machine learning rivolte all'elaborazione dei testi. La tecnica che verrà utilizzata in SAPI per costruire il classificatore sarà quella Naive Bayes.

✓ **Generazione automatica dei dialoghi vocali**

A valle delle fasi di analisi strutturale, analisi semantica e tematica, metadazione e indicizzazione di un documento digitale convenzionale lo stesso viene corredato delle seguenti informazioni ausiliarie statiche atte a migliorare, soprattutto in termini di efficienza, la fruizione del documento da parte dell'utente:

- l'insieme dei metadati descrittivi (*titolo, autore, editore, data di pubblicazione, ...*);
- l'insieme dei metadati strutturali (*capitolo, paragrafo, confine tematico, referenza bibliografica, nota, figura, ...*) che andranno a costituire una sorta di indice generale comprensivo di eventuali note e figure presenti nel documento;
- il file indice a supporto della navigazione;

Dopo alcune sessioni di consultazione del documento da parte dell'utente, in aggiunta alle precedenti informazioni statiche possono essere disponibili le seguenti informazioni ausiliarie dinamiche:

- i segnalibri (bookmarks) introdotti dall'utente;
- la predizione del comportamento dell'utente (predizione della prossima richiesta dell'utente) dedotta automaticamente dal sistema.

In aggiunta alle voci del menu che consentono di richiedere le funzionalità di base quali regolazione volume TTS, velocità di lettura, inserimento/rimozione segnalibro, etc., le suddette informazioni ausiliarie statiche e dinamiche saranno utilizzate dal modulo Vocal Dialogues Handler per preparare i dialoghi vocali; esse rappresentano, infatti, le frasi da sintetizzare e gli elementi costituenti le grammatiche statiche e dinamiche che verranno utilizzate per il riconoscimento vocale *speaker independent*.

Per ciascun documento digitale, Il Vocal Dialogues Handler, sarà in grado di generare due tipi di dialoghi vocali:

1. Dialoghi predetti; sono i dialoghi ricavati in base alle preferenze dell'utente dedotte automaticamente dallo Screen reader a partire dai dati storici relativi al comportamento dell'utente;
2. Dialoghi assoluti; sono i dialoghi completi che prescindono dalle preferenze comportamentali (implicite) dell'utente.