

tesi di laurea

Esperimenti di migrazione di funzionalità di Web Application verso Web Service

Anno Accademico 2007/2008

relatore

Ch.mo prof. Porfirio Tramontana

candidato

Fabio Conte

Matr. 534/002716

Introduzione

**Sul Web sono presenti molte Web Application che offrono diverse funzionalità.
Come possiamo renderle accessibili in maniera programmatica?**

- Una possibile soluzione è migrarle verso Web - Service

Definizione del W3C:

“ A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically: WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.”

[W3C Working Group – February 11, 2004]

Quale metodologia utilizzare per la migrazione?

Black-Box: nella quale lo studio si limita alle sole interfacce esterne del sistema e ad attività di incapsulamento (*wrapping*) costruendo dei bozzoli software intorno a quelle unità che funzionano bene ed hanno interfacce ben definite

White-Box: la quale richiede una conoscenza approfondita della Web-Application in modo da consentire una sostituzione del sistema (netta o graduale), adottando il codice in nuovi ambienti, riscrivendone alcune parti ed aggiungendone, in modo coerente, di nuove

Gray-Box: nella quale si utilizza un misto di black-box e white-box, utilizzando il secondo approccio soprattutto per le parti con alto valore di business.

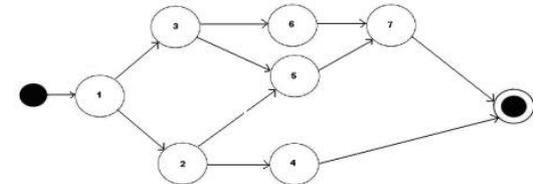
Andremo ad analizzare l'approccio black-box.

Sarebbe troppo oneroso sviluppare un wrapper ad hoc per ogni funzionalità che si vuole migrare.

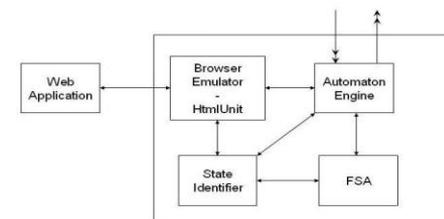
Possiamo "scomporre" il wrapper in due parti:

- 1) Automa non deterministico
- 2) Interprete

Esempio di Automa



L'interprete



Passi della metodologia black-box:

- Selezionare la funzionalità di interesse che si vuole migrare verso web service
- Effettuare un reverse engineering dell'interfaccia utente della Web Application
- Creare il Modello di Interazione descrivendolo tramite l'FSA
- *Convalidare* il Wrapper
- Esportare la funzionalità creata verso web service

Esempio Migrazione di un Servizio: Google Scholar

INPUT	Nome e Cognome
OUTPUT	Uri, Titolo, Citazioni
ECCEZIONE	ERR404 se l'automata termina in uno stato non noto ERR001 se la ricerca non produce risultati
BEHAVIOUR	Restituisce, dato il nome di una persona, tutti gli articoli (nella forma di uri, titolo e citazioni) in cui compare tra gli autori.

Home Page:



■ “Prima”

Google Scholar BETA Porfirio Tramontana Cerca Ricerca avanzata Google Scholar Preference Scholar Guida Scholar

Scholar Tutti gli articoli - [Articoli recenti](#)

[Reverse engineering Web applications: the WARE approach](#)
3A Di Lucca, AR Fasolino, P Tramontana - Journal of Software Maintenance and Evolution: Research and ..., 2004 - interscience.wiley.com
.. Research Reverse engineering Web applications: the WARE approach Giuseppe Antonio Di Lucca 1, Anna Rita Fasolino 2,* , † and **Porfirio Tramontana** 2 ...
Citato da 25 - [Articoli correlati](#) - [Ricerca Web](#) - [ACNP Pos'eduto Biblioteche](#) - [Tutte e 2 le versioni](#)

[Supporting Web application evolution by dynamic analysis](#) - ▶ [unisannio.it](#) (PDF)
GA Di Lucca, M Di Penta, AR Fasolino, P Tramontana - Principles of Software Evolution, Eighth International ..., 2005 - IEEE Explore.IEEE.org
.. Giuseppe Antonio Di Lucca*, Massimiliano Di Penta*, Anna Rita Fasolino*, **Porfirio Tramontana*** (dilucca, dipenta@unisannio.it, (fasolino, ptramont)@unina.it...
Citato da 3 - [Articoli correlati](#) - [Ricerca Web](#) - [Tutte e 7 le versioni](#)

Discriminante

Avanti

Compare tra gli autori, risultato inserito

Non Compare tra gli autori, non va inserito

Google Scholar BETA Porfirio Tramontana Cerca Ricerca avanzata Google Scholar Preference Scholar Guida Scholar

Pagina dei risultati: 1 2 3 4 5 6 7 8 9 10 Avanti

Click

Nome	“Prima”
Discriminante	//FONT[@size="-1"]][contains(., "Pagina dei risultati")] and //img/attribute::src[contains(., "/intl/it/nav_next.gif")]
Descrizione	Dobbiamo prelevare dalla pagina le informazioni di interesse. Per recuperarle utilizzeremo gli xpath. In tabella alcuni esempi. Nei risultati dobbiamo inserire solo gli articoli dell'autore oggetto di ricerca. Possiamo realizzare ciò con un semplice script interno all'automata. Per recuperare gli altri risultati dobbiamo cliccare sul link “Avanti”
Stato Successivo	“Centrale”, “Ultima”

Informazione	Xpath
Titolo1	/HTML/BODY[1]/P[1]/SPAN[1]/A[1]/text()
Uri1	/html/body/p[1]/span/a//attribute::href
Autori1	/HTML/BODY[1]/P[1]/FONT[1]/SPAN[1]/text()
Citazioni1	/html/body/p[1]/font/a[contains(., 'Citato')]/text()

Centrale



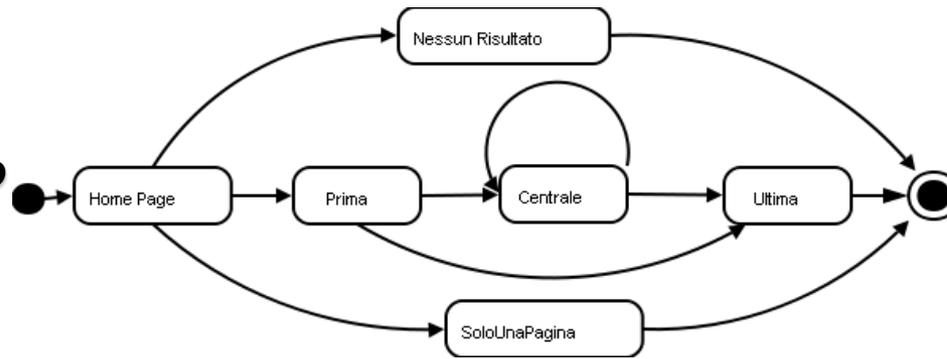
Nome	“Centrale”
Discriminante	<code>//img/attribute::src[contains(., "nav_previous")] and //img/attribute::src[contains(., "/intl/it/nav_next.gif")]</code>
Descrizione	Preleveremo le informazioni con lo stesso criterio dello stato “Prima” Una volta prelevate dovremo cliccare sul link “Avanti” per scorrere gli altri risultati.
Stato Successivo	“Centrale”, “Ultima”

Ultima

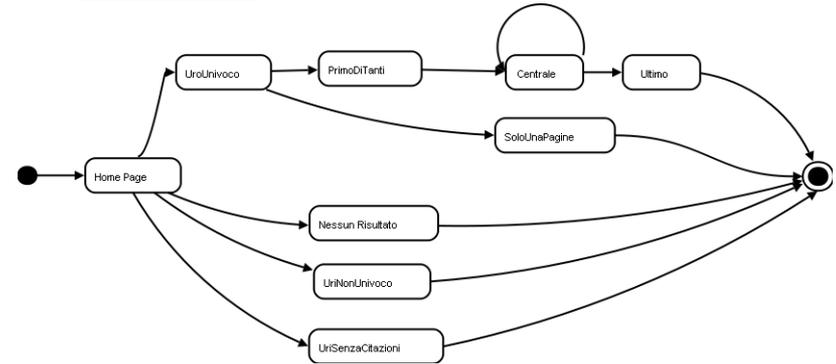


Nome	“Ultima”
Discriminante	<code>//img/attribute::src[contains(., "nav_previous")] and not(//img/attribute::src[contains(., "/intl/it/nav_next.gif")])</code>
Azioni Da Compiere	Siamo nell’ ultima pagina dei risultati. Dobbiamo prelevare nuovamente le informazioni di interesse, con lo stesso criterio di “Prima”
Stato Successivo	“Finale”

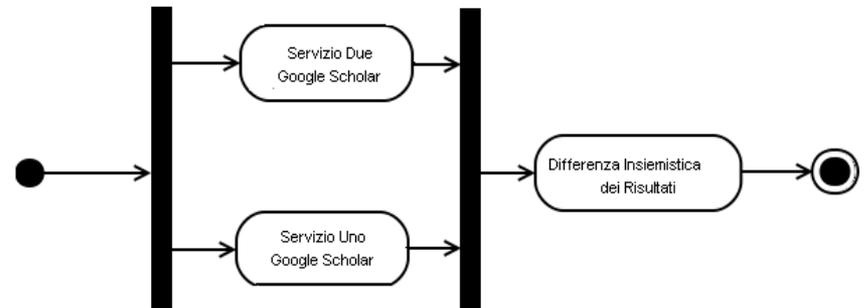
Automa per il servizio Uno



Google Scholar Servizio Due	
INPUT	Uri Articolo
OUTPUT	Uri, Titoli
ECCEZIONE	ERR404 se l'automa termina in uno stato non noto ERR001 l'articolo cercato non esiste ERR002 se l'articolo non è univoco ERR003 se non sono presenti citazioni
BEHAVIOUR	Dato un uri univoco di un articolo, esso restituisce, se esistenti tutti gli articoli che lo hanno citato (Uri e titolo).



Google Scholar Servizio Tre	
INPUT	Uri Articolo, Autore
OUTPUT	Uri, Titoli
ECCEZIONE	ERR404 se l'automa termina in uno stato non noto ERR001 l'articolo cercato non esiste ERR002 se l'articolo non è univoco
BEHAVIOUR	Dato un uri di un articolo, e un suo autore, restituisce tutte le citazioni che non siano autocitazioni

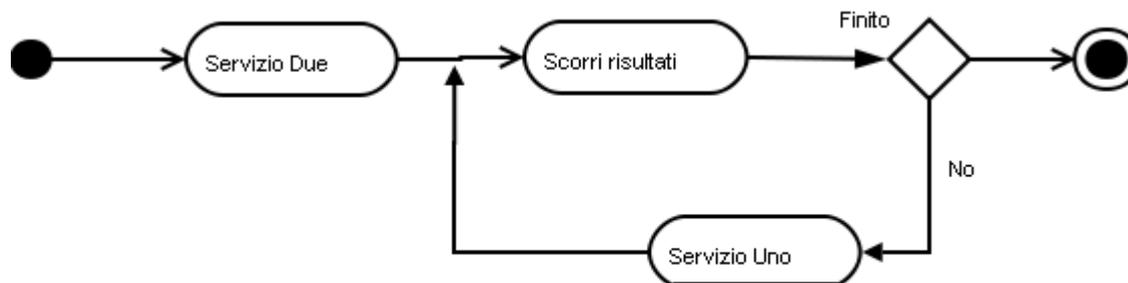
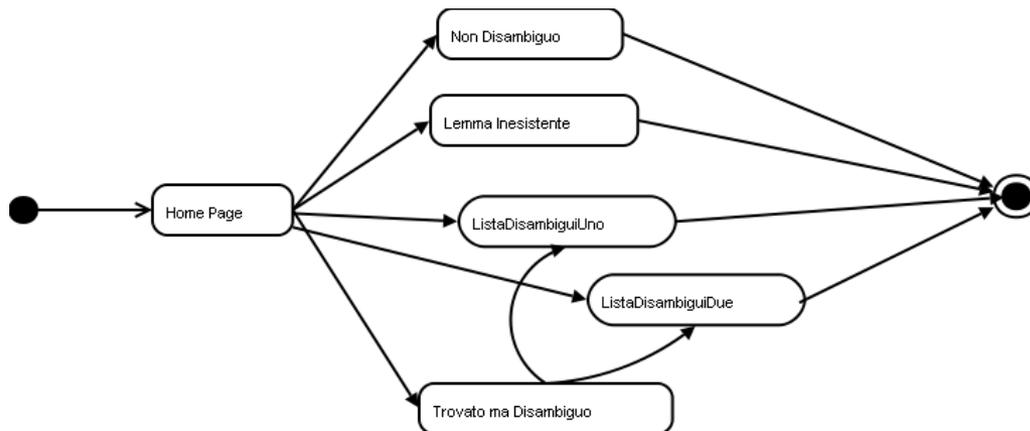
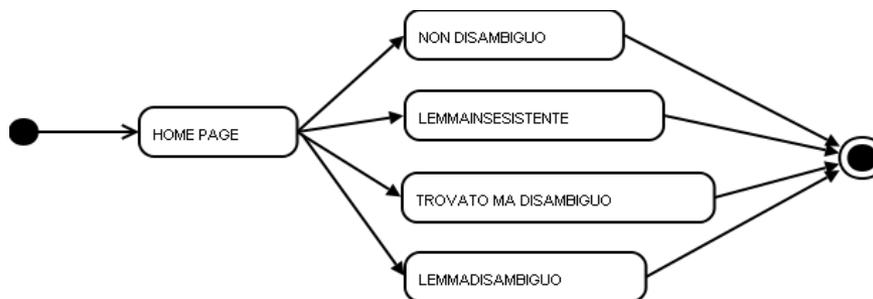


Wrapper Wikipedia

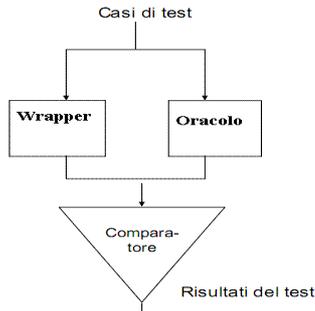
Wikipedia Servizio Uno	
INPUT	Lemma
OUTPUT	URI, Definizione, Suggerimento
ECCEZIONE	ERR404 se l'automa termina in uno stato non noto ERR001 se il lemma cercato non esiste ERR002 se il lemma cercato è ambiguo
BEHAVIOUR	Restituisce, l'incipit del lemma presente su wikipedia

Wikipedia Servizio Due	
INPUT	Lemma
OUTPUT	Lista delle ambiguità, Parole Chiave, uri
ECCEZIONE	ERR404 se l'automa termina in uno stato non noto ERR001 se il lemma cercato non esiste ERR002 se il lemma cercato non è ambiguo
BEHAVIOUR	Restituisce la lista delle possibili disambiguazioni

Wikipedia Servizio Tre	
INPUT	Lemma
OUTPUT	URI, Definizione, Suggerimento
ECCEZIONE	ERR404 se l'automa termina in uno stato non noto ERR001 se il lemma cercato non esiste ERR002 se il lemma cercato è ambiguo
BEHAVIOUR	Restituisce, l'incipit del lemma presente su wikipedia



Per convalidare i wrapper effettueremo un test funzionale



Servizio	Numero Test	Positivi	Percentuale	Tempo Esecuzione (medio)	TempoEsecuzione/ Numero di stati
Wikipedia Servizio Uno	987	986	99,89%	2064 ms	668
Wikipedia Servizio Due	987	985	99,79%	2306 ms	684
Imdb Servizio Uno	204	204	100%	5423 ms	1091,33
Imdb Servizio Due	326	326	100%	3149 ms	1089,67
Google Scholar Servizio Uno	87	87	100%		940,01
Google Scholar Servizio Due	50	50	100%		938,12

Esempi di Tempi di esecuzione dettagliati per il Servizio "Wikipedia Due"

Lemma	Ti(ms)	Te(ms)	Tr(ms)	Stato	Te(ms)	Tr(ms)	Stato	Te(ms)	Tr(ms)	Stato	Te(ms)	Tr(ms)	Stato	Ti Tempo Inizializzazione	Te Tempo Elaborazione	Tr Tempo Recupero
Apple	205,06	77	828	HomePage	95	1422	TrovatoMaDisambiguo	55	578	ListaDisambigui2	83	0	EndState			
Pomaceous	145,21	62	734	HomePage	90	921	Lemmalnesistente	52	0	EndState						
Tree	142,13	74	1156	HomePage	95	1219	TrovatoMaDisambiguo	69	625	ListaDisambigui2	50	0	EndState			
Malus	176,09	93	844	HomePage	62	1000	TrovatoMaDisambiguo	82	531	ListaDisambigui	53	0	EndState			

Il tempo di esecuzione dipende da:

- Numero di stati in cui transita l'automa
- Tempo di recupero per le pagine web associate ad ogni stato
- Tempo di elaborazione per le pagina web associate ad ogni stato

$$T_{es} = T_{ic} + \sum_{l=1}^n ((Tr_{pl} + Te_{pl}))$$

Conoscendo il numero di stati in cui transiterà l'automa, il tempo medio di recupero e di elaborazione, con buona probabilità il suo tempo di esecuzione sarà:

$$T_{es} \approx T_i + (Tr + Te) * (n - 1) + Te$$

Comparativa tra diverse metodologie:

Wikipedia Servizio Due

Metodologia	Tempo (medio)
White-Box (en.wikipedia.org)	380 ms
Gray-Box (Wikipedoa-lab.org)	890 ms
Black-Box (Le nostre)	2064 ms

Come prevedibile l'approccio black-box è il più lento. In generale le soluzioni basate sul wrapping non brillano mai per efficienza di tempo di esecuzione. Ma in alcuni casi è l'unico possibile.

- **Per migliorare le prestazioni, per servizi idempotenti possiamo salvare i risultati ottenuti dall'esecuzione del wrapper.**
Possiamo considerare idempotenti anche quei servizi dipendenti da resource, in cui si può approssimare la costanza della risorsa tra richieste ravvicinate (es: motori di ricerca)

Svantaggi:

- **Piccolo overhead introdotto ad ogni richiesta**
- **Dimensione dell'archivio non trascurabile**
- **Possibilità di incoerenza delle informazioni**

Sviluppi futuri

I risultati ottenuti sono stati incoraggianti, ma rimangono aperte una serie di problematiche legate a scelte/limiti progettuali:

- Non possiamo operare su Web Application le cui pagine Web non siano codificate in Html**
- Non possiamo operare su Web Application sviluppate secondo la metodologia AJAX**
- Gli elementi creati con le funzioni Javascript “alert” e “confirm” non sono gestibili, per limiti delle libreria utilizzata per il “Browser Emulator”**
- Si potrebbe pensare di aggiungere nuovi elementi di interazione con la Web Application, attualmente non supportati, come il doppio click o il movimento del mouse su un elemento**