

PYTHON, RESEARCH, BUSINESS AND INDUSTRY SUPPORT

Fabio Pliger¹ Elisa Burato²

¹S.I.A Castel D'Azzano, Verona

²Dip. Informatica, Università degli Studi di Verona

June 8, 2007

STRUMENTI DI SVILUPPO

- Sia opera in un settore di nicchia molto particolare dove molti dei punti seguenti sono importanti... Basta saper dosare le cose!
- Ogni settore ha un suo strumento di sviluppo predestinato. Dipende solo da esigenze di...
 - Prestazioni
 - Stabilità
 - Scalabilità
 - Costi Manutenibilità
 - Documentabilità
 - Semplicità
 - Curva di apprendimento
 - Versatilità
 - (Simpatia e Pubblicità)
 - ... ne vuoi altre?

CONTESTO S.I.A

Struttura a 4 livelli

- ① HW
- ② PLC & Progr. Dispositivi
- ③ SCADA
- ④ Sviluppo software

REPARTO SOFTWARE

- Dare Supporto e integrare i sistemi Scada / PLC
- Arrichire le informazioni
- Proteggere i dati
- Tracciare modifiche

TOOLS

Storia precedente

- Supporto nativo degli scada: C / C++ / VB / VB Script / (.NET)
- Tools di maggior diffusione: VB / VB Script / C / C++

Fattori Positivi

- Performance
- Velocità di Inizio Progetto e Facilità interfacciamento a Scada

Fattori Negativi

- alti costi
- difficoltà di aggiornamento
- difficoltà di debug
- ...
- zero flessibilità

TOOLS MIGLIORI?

- python
- dubbi iniziali:
 - prestazioni
 - interfacciamento sistemi esistenti
 - difficoltà nell'apprendimento
 - esistenza moduli per le esigenze
- speranze iniziali
 - risoluzione problemi con S.O. diversi
 - flessibilità
 - facilità nel debugging

PRIME IMPRESSIONI A CALDO

- Curva di apprendimento praticamente nulla
- Estremamente pragmatico
- Batterie incluse
- Prestazioni al di sopra delle aspettative

AL LAVORO!

Prime applicazioni:

- Historical Analysis
- Batch Report

CONSEGUENZE

Da allora python è divenuto il nostro linguaggio di sviluppo standard.

Pregi:

- tutti i risultati sono stati al di sopra delle aspettative
- ampie librerie di supporto alle applicazioni commerciali e non (gui, connessione db, freezing, documentazione, applicazioni scientifiche ...)

Aspetti Negativi

- unicode
- (freezing)
- (editor gui)

INTRODUZIONE

È nato un progetto di ricerca tra [S.I.A. srl](#) di Castel d'Azzano (Verona) e l'[Università degli studi di Verona](#).

Obiettivi del progetto sono lo studio delle problematiche aziendali nel contesto in cui S.I.A. opera e la risoluzione di tali problematiche mediante tecniche di **rappresentazione della conoscenza**.

- analisi delle problematiche aziendali

INTRODUZIONE

È nato un progetto di ricerca tra [S.I.A. srl](#) di Castel d'Azzano (Verona) e l'[Università degli studi di Verona](#).

Obiettivi del progetto sono lo studio delle problematiche aziendali nel contesto in cui S.I.A. opera e la risoluzione di tali problematiche mediante tecniche di **raccomandazione della conoscenza**.

- analisi delle problematiche aziendali
- soluzione proposta università

INTRODUZIONE

È nato un progetto di ricerca tra [S.I.A. srl](#) di Castel d'Azzano (Verona) e l'[Università degli studi di Verona](#).

Obiettivi del progetto sono lo studio delle problematiche aziendali nel contesto in cui S.I.A. opera e la risoluzione di tali problematiche mediante tecniche di **rappresentazione della conoscenza**.

- analisi delle problematiche aziendali
- soluzione proposta università
- strumenti pensati per la soluzione: python

S.I.A.: PROBLEMATICHE

I problemi che l'azienda propone all'università di risolvere sono legati a processi di:

S.I.A.: PROBLEMATICHE

I problemi che l'azienda propone all'università di risolvere sono legati a processi di:

- raccolta ed analisi dei dati
 - ① organizzazione ad alto livello mediante **gruppi di lettura**

S.I.A.: PROBLEMATICHE

I problemi che l'azienda propone all'università di risolvere sono legati a processi di:

- raccolta ed analisi dei dati
 - ① organizzazione ad alto livello mediante **gruppi di lettura**
 - ② tecniche di **correzione semiautomatica**

S.I.A.: PROBLEMATICHE

I problemi che l'azienda propone all'università di risolvere sono legati a processi di:

- raccolta ed analisi dei dati
 - ① organizzazione ad alto livello mediante **gruppi di lettura**
 - ② tecniche di **correzione semiautomatica**
 - ③ ottimizzazione della lettura sensoriale

S.I.A.: PROBLEMATICHE

I problemi che l'azienda propone all'università di risolvere sono legati a processi di:

- raccolta ed analisi dei dati
 - ① organizzazione ad alto livello mediante **gruppi di lettura**
 - ② tecniche di **correzione semiautomatica**
 - ③ ottimizzazione della lettura sensoriale
- compressione dati con metodi di **archiviazione sofisticata**

S.I.A.: TECNOLOGIE UTILIZZATE

L'area di ricerca con cui nasce il progetto con l'università è il gruppo di **Intelligenza Artificiale**(AI).

Esistono due linee di ricerca in Intelligenza Artificiale:

- ragionamento automatico
- **rappresentazione della conoscenza**

AI: RAPPRESENTAZIONE DELLA CONOSCENZA

La ricerca scientifica per i problemi legati alla rappresentazione della conoscenza, studia le tecniche mediante cui la conoscenza può essere

- **condivisa**: esistono informazioni di interesse comune ad un gruppo di utenti

AI: RAPPRESENTAZIONE DELLA CONOSCENZA

La ricerca scientifica per i problemi legati alla rappresentazione della conoscenza, studia le tecniche mediante cui la conoscenza può essere

- **condivisa**: esistono informazioni di interesse comune ad un gruppo di utenti
- **modificata**: un gruppo di utenti alimenta la conoscenza condivisa dal gruppo con nuove informazioni

AI: PROPOSTA A S.I.A.

Nel contesto S.I.A. le informazioni che costituiscono la conoscenza da condividere e modificare sono i dati raccolti ed elaborati dal software sviluppato in azienda e provenienti da sensori analogici o digitali.

AI: PROPOSTA A S.I.A.

Nel contesto S.I.A. le informazioni che costituiscono la conoscenza da condividere e modificare sono i dati raccolti ed elaborati dal software sviluppato in azienda e provenienti da sensori analogici o digitali.

Il progetto di ricerca si propone di utilizzare:

- la definizione di ontologie per la condivisione delle informazioni di interesse aziendale per poter legare i dati raccolti in base al loro **significato aziendale**
- negoziazione semantica per controllare ed organizzare i dati raccolti per dare un **senso semantico** ad ognuno di essi in modo da arricchire i gruppi di lettura

NEGOZIAZIONE SEMANTICA

Negoziazione è un termine commerciale per indicare il processo con cui le parti coinvolte nel mercato commerciale si accordano su i termini contrattuali da rispettare.

NEGOZIAZIONE SEMANTICA

Negoziazione è un termine commerciale per indicare il processo con cui le parti coinvolte nel mercato commerciale si accordano su i termini contrattuali da rispettare.

Negoziazione Semantica è il processo mediante cui si ottiene condivisione di informazione: a partire da conoscenze individuali negoziando si ottiene una conoscenza condivisa che risulta essere una mediazione tra le conoscenze iniziali.

NEGOZIAZIONE SEMANTICA

Negoziazione è un termine commerciale per indicare il processo con cui le parti coinvolte nel mercato commerciale si accordano su i termini contrattuali da rispettare.

Negoziazione Semantica è il processo mediante cui si ottiene condivisione di informazione: a partire da conoscenze individuali negoziando si ottiene una conoscenza condivisa che risulta essere una mediazione tra le conoscenze iniziali.

Esiste quasi certezza che la conoscenza condivisa non sia tanto precisa quanto quella individuale

NEGOZIAZIONE SEMANTICA 2

La negoziazione semantica si applica al processo di arricchimento del database che raccoglie tutti i dati poichè in esso sta il concetto di **condivisione**.

NEGOZIAZIONE SEMANTICA 2

La negoziazione semantica si applica al processo di arricchimento del database che raccoglie tutti i dati poichè in esso sta il concetto di **condivisione**.

Quando giunge nuova informazione al database occorre controllare che questa non intacchi l'informazione già presente rendendo difficile la lettura dei dati per alcuni utenti.

ONTOLOGIE

Una ontologia è uno strumento che supera la funzione originale di un dizionario, in quanto i termini sono non solo annotati, ma anche vincolati, nel senso che si stabiliscono delle relazioni tra di essi.

ONTOLOGIE

Una ontologia è uno strumento che supera la funzione originale di un dizionario, in quanto i termini sono non solo annotati, ma anche vincolati, nel senso che si stabiliscono delle relazioni tra di essi.

Raccoglie dei termini e a ciascun termine associa una definizione. Il gruppo di ricerca in rappresentazione della conoscenza ha sviluppato una web application che costruisce in modo automatico una ontologia ed è anche in grado di abbinare ad ogni parola dell'ontologia il contesto in cui si inserisce; dunque è in grado di collegare altre parole che spesso occorrono insieme a quella di interesse.

ONTOCHOIR: ONTOLOGIES CHOIR

Ontochoir è una web application sviluppata in python che ha le caratteristiche di

- 1 costruire una ontologia aziendale analizzando la documentazione esistente
- 2 forum di azienda utilizzato dalla applicazione per alimentare e modificare la terminologia
- 3 visualizzare una mappa grafica dei termini utilizzati in azienda con un grafo i cui nodi sono termini e sono associati mediante relazione di iponimia (termine più o meno generale di un altro) e di omocontestualità (due termini compaiono spesso insieme)

ONTOCHOIR 2

La proposta universitaria pensa di utilizzare Ontochoir per

- definire un dizionario di azienda per migliorare la comunicazione interna ed esterna

ONTOCHOIR 2

La proposta universitaria pensa di utilizzare Ontochoir per

- definire un dizionario di azienda per migliorare la comunicazione interna ed esterna
- definire una ontologia per i dati raccolti

ONTOCHOIR 2

La proposta universitaria pensa di utilizzare Ontochoir per

- definire un dizionario di azienda per migliorare la comunicazione interna ed esterna
- definire una ontologia per i dati raccolti
- analizzare la coordinazione semantica tra i dati elaborati dal software sviluppato

ONTOCHOIR 2

La proposta universitaria pensa di utilizzare Ontochoir per

- definire un dizionario di azienda per migliorare la comunicazione interna ed esterna
- definire una ontologia per i dati raccolti
- analizzare la coordinazione semantica tra i dati elaborati dal software sviluppato
- produrre gruppi di lettura dei dati in modo automatico

ONTOCHOIR E PYTHON

OntoChoir è scritto in python perchè è necessario un trattamento delle liste rapido ed efficiente.

Un' applicazione web richiede fondamentalmente velocità di esecuzione e python soddisfa tale requisito incorporando una serie di meccanismi sulle liste di estrema funzionalità e di semplicità che non richiedono grande potenza di calcolo come altri linguaggi.

Scegliendo python come linguaggio di programmazione, la scelta del web application server va su Zope.

ONTOCHOIR: PACCHETTI ZOPE

Per lo sviluppo dell'applicazione è stato necessario l'impiego di altri pacchetti tra cui:

- GroupUserFolder: per permettere la definizione di gruppi di utenti
- DocumentFolder: per lo storing ed il parsing di file specifici

ONTOCHOIR: RICERCA SU INTERNET

Ricerca su internet dei termini e parsing della risposta

```
def searchByGigablast(self, userName, sessionID):  
    f = urllib.urlopen(query)  
    s = gigablashFilter()  
    s.feed(f.read())  
    s.close()  
    primary_urllist = s.getUrlDict().values()  
    return self.getSpecificTerms(...)
```

- `gigablahFilter()`: invoca un parser XML
- `s.feed(f.read())`: costruzione di un dizionario interno popolato da tutti i collegamenti ipertestuali

ONTOCHOIR: LETTURA DEI FILE LOCALI

Lettura dei file locali caricati e parsing del testo

```
file_ids = directory.objectIds("Document File")
file_contents = ""
for docID in file_ids:
    content = directory._getOb(docID).text_content()
    file_contents += content
filebody = analyzer.nltk_analyzerCorpusSimple(file_contents)
filter = corpusFilter(filebody, "")
filter.extractSpecificWordsByAllowedTypes(meanVarianceDict)
```

- `content = directory._getOb(docID).text_content()`: parsing di ogni file
- `nltk_analyzerCorpusSimple(file_contents)`: analisi del testo
- `extractSpecificWordsByAllowedTypes(meanVarianceDict)`: estrazione dei nomi dal testo

Grazie per l'attenzione