



Corso di Ingegneria del Software a.a. 2008/2009

Progetto *DM2008-09*

OGGETTO:

Piano Attività

Specifica e progettazione software per “DM2008-09”

SINTESI DEI CONTENUTI:

Il Piano si riferisce alle attività di progetto da effettuare nell’ambito del corso di Ingegneria del Software.

Il piano contiene le seguenti linee di attività:

- A. Analisi e specifica dei requisiti mediante notazione formale
- B. Definizione dell’architettura e progettazione del sistema
- C. Definizione di un piano di testing

Protocollo: PSE-345-2008

Data di emissione: 9/12/2008

Eseguito da: A. Staiano

Approvato da: A. Staiano

DESTINATARI:

A: Studenti di
Ingegneria del Software 2008- 2009

1. OBIETTIVI

La società *Parthenope Software Engineering* ha l’obiettivo di sviluppare un sistema software necessario per coadiuvare il lavoro di analisi di dati di un gruppo di ricercatori di un osservatorio astronomico.

Per la realizzazione del suddetto obiettivo, *Parthenope Software Engineering* è impegnata nella progettazione dei moduli software basati su metodi statistici e di apprendimento automatico necessari per svolgere alcune attività di Data Mining.

Il presente documento si riferisce alle seguenti linee di attività:

- Formalizzazione dei requisiti mediante notazione UML
- Definizione dell’architettura e progettazione del sistema, mediante notazione UML
- Definizione di un piano di testing

Per ciascuna linea di attività sono definiti gli obiettivi descritti di seguito.

1.1 Analisi e specifica dei requisiti

Scopo dell’attività è la definizione, attraverso modelli formali, delle funzionalità del sistema. In particolare si richiede la fornitura di un documento che formalizzi, attraverso le tabelle per scenari e casi d’uso e notazioni UML, tutte le funzionalità del sistema.

Per le funzionalità da progettare si rimanda al capitolo 2.

1.2 Definizione dell’architettura e progettazione del sistema

Obiettivo di questa linea di attività è la fornitura di un documento che contenga la progettazione dettagliata del sistema e di tutte le funzionalità offerte, attraverso modelli e notazioni UML

1.3 Definizione di un piano di testing

Si richiede la realizzazione di un documento che contenga un elenco di test cases per verificare e validare il sistema progettato.

2. SPECIFICHE DI PRODOTTO

Il software *DM200809* rappresenta uno strumento integrato attraverso il quale l’utente ha, tra le altre cose, la possibilità di svolgere alcune attività di Data Mining (analisi statistiche dei dati, classificazione, clustering, regressione e visualizzazione).

Il sistema deve essere organizzato mediante un insieme di interfacce utente grafiche (GUI) attraverso le quali l’utente ha la possibilità di svolgere tre attività principali:

1. Caricamento di un data set oggetto di studio e analisi statistiche preliminari,
2. Impiego di tecniche di apprendimento automatico supervisionate e non supervisionate per il clustering, la classificazione e la regressione,
3. Visualizzazione dei dati.

Per quanto concerne la prima funzionalità, l’utente deve avere la possibilità di caricare un data set il cui formato d’origine è di tipo testuale. Il sistema deve presentare un primo insieme di informazioni sui dati caricati: numero di record (detti anche oggetti o pattern), numero di campi (detti parametri o caratteristiche), numero di campi con valori nulli (NaN). In aggiunta, il sistema deve visualizzare un insieme di informazioni di tipo statistico per ogni campo del data set oggetto di studio: media, deviazione standard, valore min, valore max, numero di NaN. In questo contesto l’utente deve avere la possibilità di selezionare l’intero insieme di record/campi o un loro sottoinsieme per visualizzare i dati in grafici il cui tipo è scelto dall’utente da un insieme predefinito. Infine, l’utente deve avere la possibilità di effettuare operazioni di estrazione delle caratteristiche attraverso l’impiego della tecnica nota come Principal Component Analysis (PCA) attraverso cui scegliere l’insieme più significativo delle caratteristiche. Il risultato della PCA deve essere eventualmente salvato su disco.

Per la seconda funzionalità, l’utente deve avere la possibilità di scegliere il tipo di attività da svolgere (classificazione, regressione, clustering). Una volta scelta l’attività, l’utente deve scegliere il modello di apprendimento automatico che ritiene consono, a partire da un insieme predefinito di possibilità:

1. Classificazione
 - 1.1 Rete neurale MLP
 - 1.2 Support Vector Machine (SVM)
2. Regressione
 - 2.1 Rete Neurale MLP
 - 2.2 Metodi kernel
3. Clustering
 - 3.1 Hard e fuzzy c-means
 - 3.2 Clustering gerarchico

Ciascun modello elencato nelle tre categorie prevede un insieme prefissato di parametri di input che l’utente deve opportunamente fornire mediante un’ adeguata interfaccia grafica. Dalla stessa interfaccia, l’utente deve essere in grado di avviare la fase di apprendimento. Il modello addestrato deve essere salvato in un file, in modo da poter essere caricato

successivamente. Per le sole operazioni di classificazione e regressione, l’utente deve avviare la fase di valutazione delle prestazioni (testing) del modello.

Infine, per ciò che concerne la terza funzionalità, l’utente deve avere la possibilità di scegliere tra differenti metodi di visualizzazione in due e tre dimensioni:

- a) Generative Topographic Mapping (GTM)
- b) Probabilistic Principal Surfaces (PPS)
- c) Self Organizing Maps (SOM)
- d) Multi Dimensional Scaling (MDS)

Anche in questo caso, una volta scelto il modello da adottare è necessario, attraverso un’opportuna GUI, inserire i parametri del modello, avviare il processo di apprendimento e visualizzare i dati con uno dei possibili grafici messi a disposizione da ciascuno di essi.

Si richiede di astrarre il design per favorire il riutilizzo del codice e la futura implementazione di altre funzionalità.

Per lo svolgimento della suddetta attività è richiesto l’utilizzo di un tool di CASE, scelto tra i seguenti:

- Poseidon
- ArgoUML
- DIA
- VisualParadigm (Community Edition)

3. OUTPUT ATTESI DAL CLIENTE

Per le attività sopracitate vanno realizzati i seguenti output:

1. Documento dei Requisiti Software
 - 1.i Modello Funzionale.
 - a. Specifiche dei Requisiti del Sistema.
 - b. Modellazione dei casi d'uso.
 - c. Mock-up interfaccia utente.
 - d. Una proposta di impegno risorse e di pianificazione dettagliata dell'attività, con diagrammi di Gantt.
 - e. Glossario.
 - 1.ii Modello ad oggetti di analisi e modello dinamico.
 - f. Classi, oggetti e relazioni di analisi.
 - g. Analisi dell'architettura.
 - h. Tabelle di descrizione per ogni caso d'uso.
 - i. Diagrammi di sequenza (solo per i casi d'uso che si ritiene opportuno).
 - j. Diagrammi di stato.
2. Documento di progettazione del sistema.
 - a. Diagramma delle classi (Classi di progettazione).
 - i. Rifinitura delle relazioni di analisi.
 - b. Definizione del sistema e dei suoi sottosistemi.
 - c. Diagrammi di stato.
 - d. Diagrammi di sequenza.
3. Documento di piano di testing del sistema.
 - a. Test Plan.

4. TEMPISTICHE DEGLI OUTPUT

Nella tabella sottostante sono dettagliate le principali milestone per il completamento delle attività su indicate.

<i>Fornitura</i>	<i>Output Atteso</i>	<i>Tempistica delivery</i>
Analisi e specifica dei requisiti	Documento di analisi e specifica dei requisiti	20 giorni prima dell'appello
Definizione dell'architettura e progettazione del sistema	Documento di progettazione del sistema	10 giorni prima dell'appello
Definizione di un piano di testing	Documento di piano di testing del sistema.	5 giorni prima dell'appello

5. PROCESSO DI COMUNICAZIONE

Il processo di comunicazione avverrà principalmente attraverso riunioni da effettuarsi presso lo studio del Committente (il prof. Staiano “astronomo”), previo appuntamento via e-mail, all’indirizzo antonino.staiano@uniparthenope.it.

In tali riunioni, si richiede l’utilizzo di modelli formali per la comunicazione.



6. MODALITÀ DI CONTROLLO

Il controllo della qualità del lavoro svolto verrà effettuato in corrispondenza delle consegne durante le quali il Committente valuterà la completezza ed i contenuti del risultato raggiunto nonché il superamento delle consegne di riferimento, attraverso feed-back diretti.

