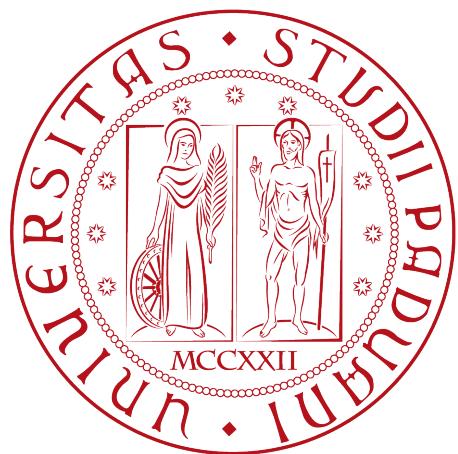


Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA

CORSO DI LAUREA IN INFORMATICA



Tres - Tree recommendation system

Tesi di laurea triennale

Relatore

Prof. Tullio Vardanega

Laureando

Alberto Andeliero

ANNO ACCADEMICO 2015-2016

Alberto Andeliero : *Tres - Tree recommendation system*, Tesi di laurea triennale, ©
Jan 2016.

Sommario

Il presente documento descrive il lavoro svolto durante il periodo di stage, della durata di circa trecento ore, dal laureando Alberto Andeliero presso l'azienda Nextep S.r.l. a Carmignano Di Brenta(PD).

Indice

1 L'azienda	1
1.1 Presentazione azienda	1
1.1.1 L'azienda	1
1.1.2 Prodotti e servizi	2
1.1.3 Tecnologie di riferimento	3
1.2 Processi aziendali	5
1.2.1 Metodologia agile	5
1.3 Strumenti a supporto dei processi	7
1.3.1 Gestione di progetto	7
1.3.2 Documentazione	8
1.3.3 Sistema di versionamento	8
1.3.4 Ambiente di sviluppo	8
1.3.5 Sistemi operativi	9
1.4 Clientela tipo e propensione all'innovazione	9
2 Il progetto nella strategia aziendale	11
2.1 L'azienda e gli stage	11
2.2 Il progetto	12
2.2.1 Il progetto	12
2.2.2 Obiettivi	12
2.3 Vincoli	13
2.3.1 Vincoli temporali	13
2.3.2 Vincoli metodologici	13
2.3.3 Vincoli tecnologici	15
2.4 Obiettivi personali	17
3 Il progetto di stage	19
3.1 Organizzazione	19
3.1.1 Pianificazione	19
3.2 Analisi dei requisiti	20
3.2.1 Requisiti	20
3.2.2 Casi d'uso	22
3.3 Progettazione architetturale	23
3.3.1 Introduzione	23
3.4 Progettazione di dettaglio	24
3.5 Verifica e Validazione	24
3.5.1 Resoconto risultati	24

4 Valutazione Retrospettiva	27
4.1 Bilancio sui risultati	27
4.1.1 Obiettivi conseguiti	27
4.1.2 Obiettivi personali	27
4.2 Bilancio formativo	27
4.3 Distanza tra formazione universitaria e lavoro	27
A Appendice A	29
Bibliografia	31

Elenco delle figure

1.1	Logo di Nextep srl	1
1.2	Sito Associazione Maestri Sci Italiani: realizzato da Nextep. Immagine tratta da Nextep	2
1.3	Classificazione delle tecnologie utilizzate	3
1.4	Diagramma della metodologia di sviluppo <i>agile</i> . Immagine tratta da GlobalTeckz	5
1.5	Diagramma del modello <i>Scrum</i> . Immagine tratta da Wikipedia	7
1.6	Piattaforma di gestione di progetto Jira	8
1.7	Classificazione degli strumenti utilizzati	9
2.1	Esempio di un albero decisionale	12
2.2	Pannello di controllo travis	14
2.3	Tipica architettura di un'applicazione che utilizza Play Framework. Immagine tratta da Omer Haderi Blog	15
2.4	Esempio di modellazione in OrientDb. Immagine tratta da Scalac Blog	17
3.1	UC0: Tres	22
3.2	UC1: Tres	23
3.3	Visione generale architettura <i>client-server</i> . Immagine tratta da Content Deliverance	24
3.4	Visione generale dell'architettura	24

Elenco delle tabelle

3.1	Tabella dei requisiti	22
3.2	Tabella dei test di unità	25
3.3	Tabella dei test di integrazione	26

Capitolo 1

L’azienda

Questo capitolo tratta dettagliatamente della azienda ospitante, il suo business, l’organizzazione interna e l’innovazione.

1.1 Presentazione azienda

Questa sezione descrive l’azienda che ha ospitato lo *stage*, elencando generalità e la struttura generale dell’azienda. Inoltre si illustrano alcuni prodotti e i servizi offerti da Nextep srl alla propria clientela.

1.1.1 L’azienda

L’azienda con cui ho svolto lo stage formativo è Nextep srl¹, la cui sede operativa si trova a Carmignano Di Brenta(PD). Nextep è una società fondata nel 2000 che opera nel settore dell’informatica e della comunicazione. Il *core business* dell’azienda è riconosciuto nelle attività di progettazione, realizzazione e gestione di infrastrutture di *Information and Digital Communication Technology*. L’organico di Nextep srl è formato da una ventina di persone, suddivise tra *web marketing* e tecnici(sviluppatori, grafici e sistemisti). L’ambiente di lavoro è condiviso con altre due aziende: Allos e Zero12. Allos si occupa della formazione del capitale umano di organizzazioni medio grandi mentre Zero12 si occupa di servizi *cloud* e applicazioni *mobile*. Il risultato è un ambiente di *co-working*, dove le competenze del singolo vengono messe a disposizione di tutti, in modo da facilitare la crescita professionale dei dipendenti delle varie aziende. Non è raro che le tre aziende collaborano tra di loro per la realizzazione di alcuni progetti.



figura 1.1: Logo di Nextep srl

¹<http://www.nextep.it/>.

1.1.2 Prodotti e servizi

Il principale settore di attività dell’azienda è il *Web*. Nextep è specialista nella progettazione e sviluppo di siti *web*, portali, *social intranet* e soluzioni di *e-commerce*. L’azienda fornisce sistemi per il *knowledge management*, che sono dei sistemi per raccogliere, sviluppare, conservare e rendere accessibile la conoscenza delle persone che fanno parte di una organizzazione. Inoltre si occupa di analisi, progettazione e realizzazione di applicazioni mobile. L’azienda pone il suo focus nella realizzazione di sistemi con interfacce innovative e soprattutto accessibili all’utente grazie alle competenze acquisite nel ramo della *User Interface* e *User Experience*. Realizza sistemi per l’integrazione con tecnologie e servizi di *Cloud Computing*.

I principali servizi offerti da Nextep srl ai propri clienti sono:

- Fornitura di servizi a canone (*server virtuali*, applicazioni *web*, servizi di *monitoring* di infrastrutture *Information and Communications Technology (ICT)*, *backup-online*, canoni di assistenza e manutenzione dei sistemi, canoni di gestione dei sistemi);
- Elaborazione dati, analisi e supporto decisionale in ambito *web marketing*, *social marketing* e *social intranet*;
- Servizio Clienti, attività di supporto tecnico e di gestione di infrastrutture *ICT* e siti *web*.

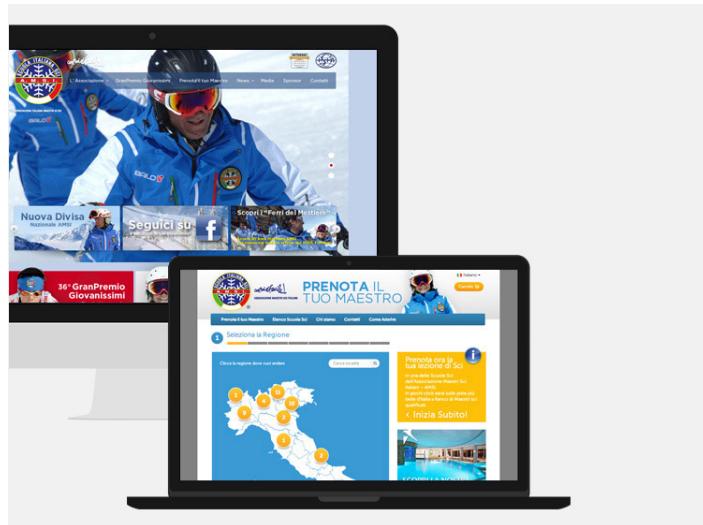


figura 1.2: Sito Associazione Maestri Sci Italiani: realizzato da Nextep. Immagine tratta da [Nextep](#)

In particolare l’azienda fornisce servizi per migliorare l’efficacia delle strategie di comunicazione *web* dedicando particolare attenzione alla reputazione e all’identità digitale. Collabora, inoltre, assieme alle aziende nella gestione delle informazioni digitali, della sicurezza e della disponibilità dei dati e delle applicazioni. Nel ramo del *marketing*, Nextep realizza azioni di *web marketing* e fornisce consulenza *web marketing*, *social marketing* e *web analytics*.

1.1.3 Tecnologie di riferimento

Essendo il *web* un settore soggetto a continue evoluzioni, le tecnologie utilizzate cambiano col tempo. Nextep cerca di aggiornare le tecnologie utilizzate in modo da fornire ai propri clienti prodotti sempre migliori e innovativi.

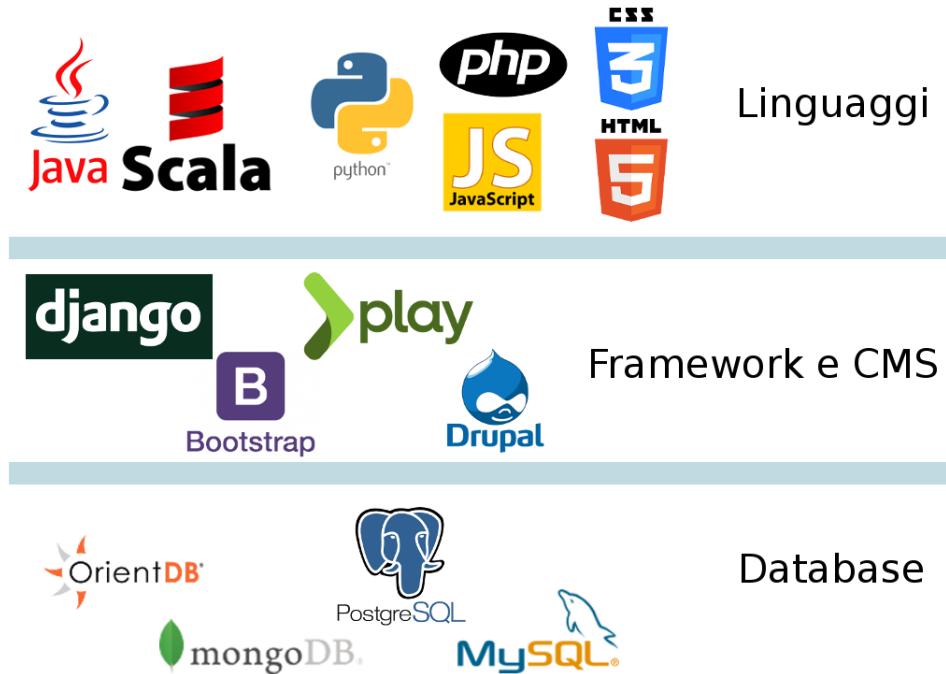


figura 1.3: Classificazione delle tecnologie utilizzate

Linguaggi di programmazione

L’azienda lavora principalmente in ambito *web*, quindi utilizza i linguaggi di programmazione più moderni per lo sviluppo dei prodotti. I linguaggi di programmazione più utilizzati in ordine di importanza sono Php, Python, Java e Scala. Per quanto riguarda le interfacce grafiche naturalmente utilizza tecnologie quali Html5, Css3 e Javascript.

Framework e CMS

Django è un *web framework open source* per lo sviluppo di applicazioni *web*, scritto in linguaggio Python, seguendo il *pattern model-view-controller*. Fornisce nativamente delle applicazioni per la gestione dei contenuti e degli accessi.

Play Framework è un *web framework open source* scritto in Java e Scala, che implementa il *pattern model-view-controller*. Il suo scopo è di migliorare la produttività degli sviluppatori usando il paradigma *Convention Over Configuration*.

Bootstrap è *framework* per la creazione di siti e applicazioni per il *web*. Essa contiene modelli di progettazione basati su Html e Css, sia per la tipografia, che per le varie componenti dell'interfaccia come moduli, bottoni e navigazione, e altri componenti dell'interfaccia.

Drupal è una piattaforma *software* di *Content Management System (CMS)*, modulare, scritta in linguaggio Php e distribuita sotto licenza GNU GPL. Grazie alla sua architettura modulare permette il riutilizzo sistematico del codice scritto e quindi un vantaggio in termini di tempo e manutenzione.

Database

Il *team* di sviluppo utilizza per la persistenza sia *database* relazionali sia non-relazionali. Fra i *database* relazionali impiega per la maggiore le tecnologie quali MySql, Sql Server e PostgreSql. Per quanto riguarda i *database* non-relazionali utilizza soluzioni come MongoDb oppure OrientDb. A seconda del contesto vengono studiati i pro e contro, in modo da poter trarre tutti i vantaggi possibili.

1.2 Processi aziendali

Questa sezione illustra l'organizzazione interna dell'azienda, descritta per processi.

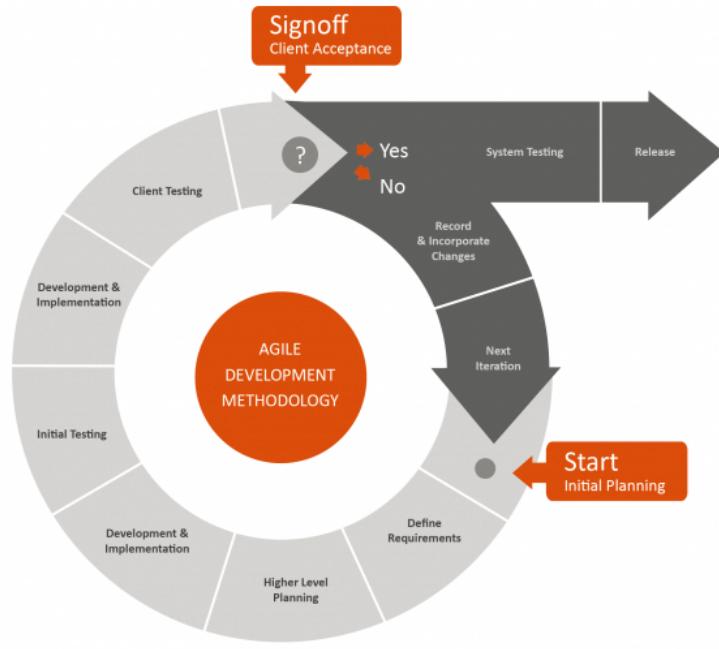


figura 1.4: Diagramma della metodologia di sviluppo *agile*. Immagine tratta da [GlobalTeckz](#).

1.2.1 Metodologia agile

L'azienda per la gestione dei progetti adotta la metodologia *agile*². Questa metodologia si basa su 4 principi:

- Gli individui e le interazioni più che i processi e gli strumenti;
- Il *software* funzionante più che la documentazione esaustiva;
- La collaborazione col cliente più che la negoziazione dei contratti;
- Rispondere al cambiamento più che seguire un piano.

In particolare tra le metodologie *agile* l'azienda segue nello specifico il modello *Scrum* dove pianifica ciclicamente degli *sprint*(un periodo di 2-4 settimane).

²<http://www.agilemanifesto.org/>.

Gli elementi costruttivi principali in *Scrum* sono:

- **Product Backlog**, è un documento ad alto livello per l'intero progetto. Contiene una lista delle funzionalità desiderate con priorità assegnate in base al valore di *business* e i *Backlog Item* che sono descrizioni di tutte le caratteristiche richieste;
- **Sprint Backlog**, è un documento che contiene informazioni su come il *team* sta procedendo nell'implementare le funzionalità per lo *sprint* successivo. Le funzionalità vengono prelevate dalla pila del documento di *Product Backlog*, in una quantità fattibile per essere realizzata nello *sprint*;
- **Burn Down Chart**, è un grafico che rappresenta il lavoro nello *Sprint Backlog* che deve essere ancora completato. Aggiornato ogni giorno, dà una semplice visione di come procede lo *sprint*;
- **Incremento**, è la somma di tutti gli elementi del *Product Backlog* completati durante uno *sprint* e tutti gli *sprint* precedenti.

Le principali pratiche che Nextep svolge sono:

- **Daily scrum**, detto anche *stand up*, è una riunione che tiene quotidianamente e serve per verificare lo stato di avanzamento durante uno *sprint*;
- **Scrum of Scrums**, è una riunione che tiene meno frequentemente del *Daily Scrum* e consente ai *team* di discutere del loro lavoro, concentrandosi su aree di sovrapposizione e d'integrazione;
- **Sprint Planning Meeting**, è una riunione che l'azienda svolge all'inizio di ogni *sprint*. Vengono pianificati gli obiettivi con relative priorità e le tempistiche. In questa riunione partecipano *Product Owner*, *Scrum Master*, intero *Scrum Team* e tutti i *manager* del caso interessati o dai rappresentanti della clientela;
- **Sprint Review**, al termine dello *sprint* l'azienda svolge questa riunione per ispezionare l'incremento e adattare, se necessario, il *Product Backlog*. In conformità a questo e dei cambiamenti al *Product Backlog* fatti durante lo *sprint*, i partecipanti pianificano il prossimo incremento;
- **Sprint Retrospective Meeting**, si svolge dopo la *Sprint Review*, in questa occasione lo *Scrum Team* ispeziona se stesso e crea un piano di miglioramento da attuare durante il successivo *sprint*.

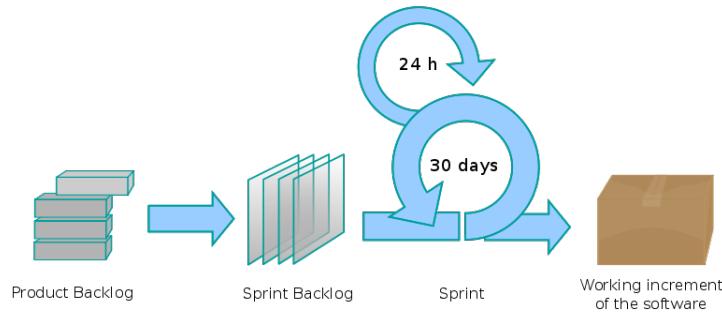


figura 1.5: Diagramma del modello *Scrum*. Immagine tratta da [Wikipedia](#).

Questa metodologia consente e promuove un maggiore coinvolgimento del cliente rispetto a quelle tradizionali, inoltre consente una maggiore flessibilità e reattività al cambio dei requisiti che possono avvenire in corso d'opera.

1.3 Strumenti a supporto dei processi

Questa sezione illustra gli strumenti utilizzati a supporto dei processi e le tecnologie utilizzate per lo sviluppo.

1.3.1 Gestione di progetto

Per fornire assistenza ai clienti l'azienda utilizza Zendesk³, che è uno strumento che permette di centralizzare tutti i canali di comunicazione tra i quali *social network*, *email* e telefono. Inoltre fornisce una panoramica generale della prestazione dell'assistenza e della soddisfazione del cliente.

Per la gestione di progetto invece utilizza lo strumento Jira⁴, che fornisce degli strumenti per la gestione di sviluppo agile con il modello *Scrum* utilizzato nell'azienda. Nello specifico, Jira permette:

- Assegnazione dei *ticket* tra i membri del *team* di sviluppo;
- Segnalazione di *issue*;
- Integrazione con il sistema di *repository*;
- Pianificazione dello *sprint*.

Nextep non utilizza la *Kanban Board* integrata in Jira, perché questo strumento di processo non è compatibile con la metodologia *agile* in uso, pertanto adopera una *Scrum Board* fisica presente in ufficio.

³<https://www.zendesk.it/>.

⁴<https://www.atlassian.com/software/jira>.

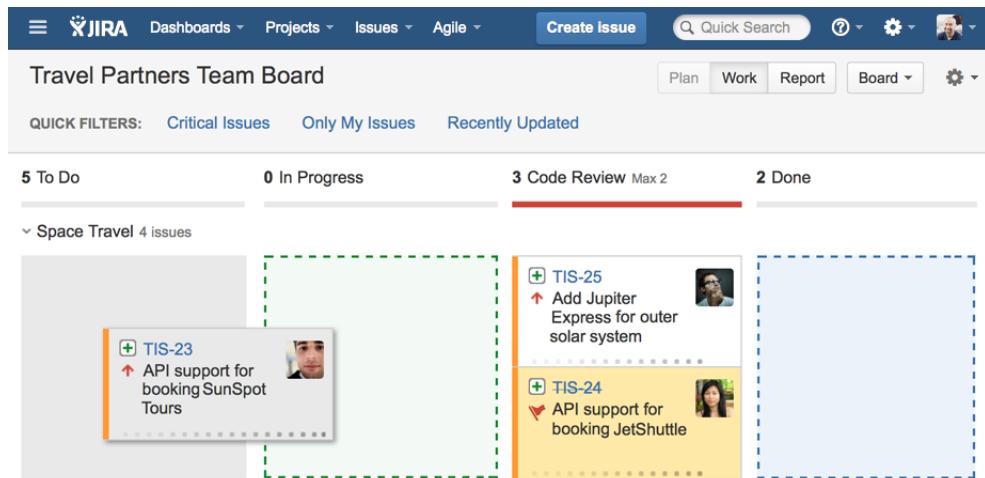


figura 1.6: Piattaforma di gestione di progetto Jira

1.3.2 Documentazione

Nextep per la produzione documentale utilizza una serie di strumenti di Google Docs, per una serie di fattori tra i quali:

- Compatibilità con tutti i sistemi operativi utilizzati;
- Permette di lavorare collaborativamente ai documenti;
- Disponibilità;
- Affidabilità.

1.3.3 Sistema di versionamento

Come sistema di versionamento il *team* utilizza lo strumento di Git⁵. I *repository* sono gestiti in un *server* locale ed occasionalmente in server esterni.

1.3.4 Ambiente di sviluppo

A seconda del contesto, il *team* utilizza l'ambiente di sviluppo che meglio si adatta alle tecnologie in uso nel progetto. In particolar modo adopera i seguenti *IDE*:

- Eclipse o IntelliJ: per lo sviluppo di applicazioni in Java e/o Scala;
- Smultron/PhpStorm: per lo sviluppo di applicazioni in Php;
- Sublime Text: per lo sviluppo in Html/Css, Javascript e Php.

⁵<https://git-scm.com/>.

1.3.5 Sistemi operativi

L'azienda non impone nessun obbligo di utilizzo di un sistema operativo. I dipendenti utilizzano, a seconda delle preferenze, sistemi operativi quali Windows, Linux e MacOs. In particolar modo quest'ultimo, visto il recente investimento aziendale in terminali MacOs, per incentivare i dipendenti alla migrazione a questo sistema operativo.



figura 1.7: Classificazione degli strumenti utilizzati

1.4 Clientela tipo e propensione all'innovazione

La tipologia di clientela per cui l'azienda lavora copre una gamma che va dal singolo privato alle organizzazioni di grandi dimensioni.

Essendo gli *e-commerce* il *core business* dell'azienda, Nextep lavora principalmente per aziende dell'ambito manifatturiero e rivendite.

L'azienda lavora per clienti/aziende italiane e molto spesso per clienti della zona visto il riconoscimento guadagnato nel padovano.

Nextep riconosce questi valori guida da seguire:

- Attitudine all'innovazione per mantenere competitività;
- Condivisione del *know how* con le altre aziende in sede.

Questi principi rendono possibile un ambiente di *co-working* dove le 3 aziende della sede sperimentano e innovano grazie alle conoscenze messe in condivisione. Per alimentare questa dinamica l'azienda promuove periodicamente la formazione dei dipendenti, in modo che essi possano essere motivati e partecipi in questo ambiente. Visto l'importanza delle dinamiche personali, l'azienda organizza ogni volta che sia necessario degli incontri di *team building* per avere sempre dei dipendenti in grado di lavorare in gruppo.

Capitolo 2

Il progetto nella strategia aziendale

Questo capitolo fornisce una descrizione dettagliata del progetto di stage, e dei motivi che hanno spinto l'azienda a proporlo oltre che i benefici derivati da tale progetto. Vengono inoltre discussi i vincoli tecnologici e metodologici imposti.

2.1 L'azienda e gli stage

Questa sezione descrive il rapporto dell'azienda con gli stage e spiega quali sono i motivi che portano l'azienda a farli, in particolare per il progetto svolto.

Data la forte evoluzione del settore *web*, l'azienda persegue l'obiettivo di esplorare i nuovi segmenti del campo per mantenere un ruolo di rilievo nel mercato. Per questo motivo c'è l'esigenza di sperimentare continuamente nuove soluzioni, per valutare possibili impieghi e vantaggi. Si vengono quindi a creare le condizioni per un'opportunità di *stage*, dove l'azienda richiede allo stagista una soluzione per le funzionalità richieste. Inoltre l'azienda sfrutta l'occasione degli *stage* per valutare una possibile collaborazione e successivamente una assunzione in azienda nel caso che il soggetto sia valutato positivamente. Nel precedente *stage*, visto l'interesse nel campo delle raccomandazioni, sono stati sviluppati dei prototipi che potessero soddisfare una raccomandazione utente in ambito *e-commerce*. I sistemi di raccomandazione in ambito commerciale sono molto utili, perché permettono un guadagno dal punto di vista economico, e dal punto di vista della *user experience* perché permettono di inserire pubblicità più pertinenti ai gusti dell'utente. In precedenza durante un recente *stage*, l'azienda aveva sviluppato il modulo di raccomandazione denominato DRE, un sistema basato sul *Collaborative filtering* che permette di fornire una raccomandazione in base alla somiglianza degli utenti.

2.2 Il progetto

Questa sezione descrive il progetto di *stage* e gli obiettivi che l'azienda ha fissato in relazione alle sue specifiche esigenze di mercato.

2.2.1 Il progetto

Il progetto di stage consiste nella realizzazione di un sistema di raccomandazione basato sulla classificazione del comportamento degli utenti all'interno del sito. Il sistema deve poter permettere l'inserimento delle azioni dell'utente ogni volta che il suo comportamento ha prodotto un risultato. Una volta collezionata una considerevole mole di dati relativi ai comportamenti, il sistema deve poter fornire una raccomandazione all'utente sulla base del suo comportamento attuale. Il comportamento di un utente è descritto tramite una serie di iterazioni e il risultato raggiunto. A sua volta una iterazione è descritta tramite l'azione e l'oggetto su cui viene svolta. Questo comportamento viene poi elaborato dall'albero di decisione generato dai dati collezionati precedentemente. La generazione dell'albero si basa sull'algoritmo di classificazione ID3 (*Iterative Dichotomiser 3*) che permette di selezionare l'attributo per discriminare il comportamento con il miglior guadagno informativo. Il sistema deve poter fornire una raccomandazione anche probabilistica nel caso di dati parziali sul suo comportamento.

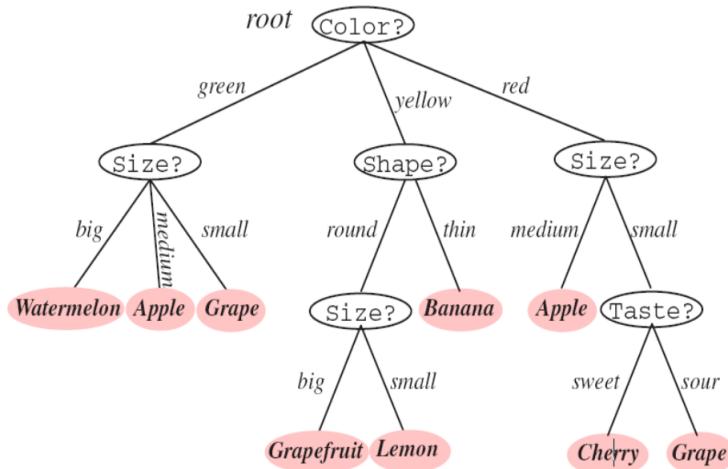


figura 2.1: Esempio di un albero decisionale

2.2.2 Obiettivi

Prima dell'inizio dello stage, Nextep ha redatto un *Piano di Lavoro* contenente gli obiettivi da realizzare nelle settimane di *stage*. Gli obiettivi erano suddivisi in minimi e massimi, successivamente in seguito ad una variazione delle priorità aziendali sono stati effettuati dei cambiamenti. Inizialmente, ho effettuato il *porting* del modulo preesistente DRE, in modo che la persistenza fosse compatibile con OrientDb e che fosse compatibile con l'ultima versione di Play Framework. Visto il protrarsi delle attività di *porting*, l'azienda ha deciso di sospendere il lavoro per poter lavorare sul nuovo modulo *Tres*.

Qui di seguito presento gli obiettivi dello stage:

- Obiettivi formativi

- Formazione sui sistemi di raccomandazione e degli algoritmi di apprendimento;
 - Studio delle tecnologie utilizzate, Scala e OrientDb;
 - Studio degli strumenti utilizzati, in particolar modo dei *framework* e degli *IDE*.

- Obiettivi minimi

- Utilizzo del database OrientDb.

- Obiettivi massimi

- Implementazione di nuovi modelli di algoritmi di apprendimento per migliorare l'individuazione dei gusti dell'utente;
 - Migliorare la fase di apprendimento del modello stesso;
 - Implementazioni di algoritmi di *cluster* per il raggruppamento di elementi omogenei in un insieme di dati.

2.3 Vincoli

Questa sezione elenca i vincoli relativi al progetto di *stage*.

2.3.1 Vincoli temporali

Per lo *stage* i vincoli temporali sono imposti direttamente dall'università, essa stabilisce una durata minima di 300 ore e una durata massima di 320 ore. Per questo motivo ho modulato il *Piano di Lavoro* in modo da distribuire le ore lavorative in 8 settimane. Con l'azienda ho concordato un impegno di 300 ore, divise in 37.5 ore settimanali. Le variazioni del piano sono state concordate con il *tutor* aziendale in maniera da rispettare i vincoli temporali imposti in seguito a ritardi dovuti al *porting* del modulo DRE.

2.3.2 Vincoli metodologici

Come strumento di versionamento ho scelto Git perché già utilizzato in precedenza e per il supporto che riceve da molti strumenti di sviluppo. Per la documentazione del codice ho scelto Scaladoc perché viene fornito insieme al compilatore e similmente a Javadoc permette di documentare direttamente nel codice sorgente. Per il *testing* nel progetto ho utilizzato TravisCI, esso fornisce un servizio di *continuous integration* facilmente integrabile con i *repository* di Github¹.

¹<https://github.com/>.

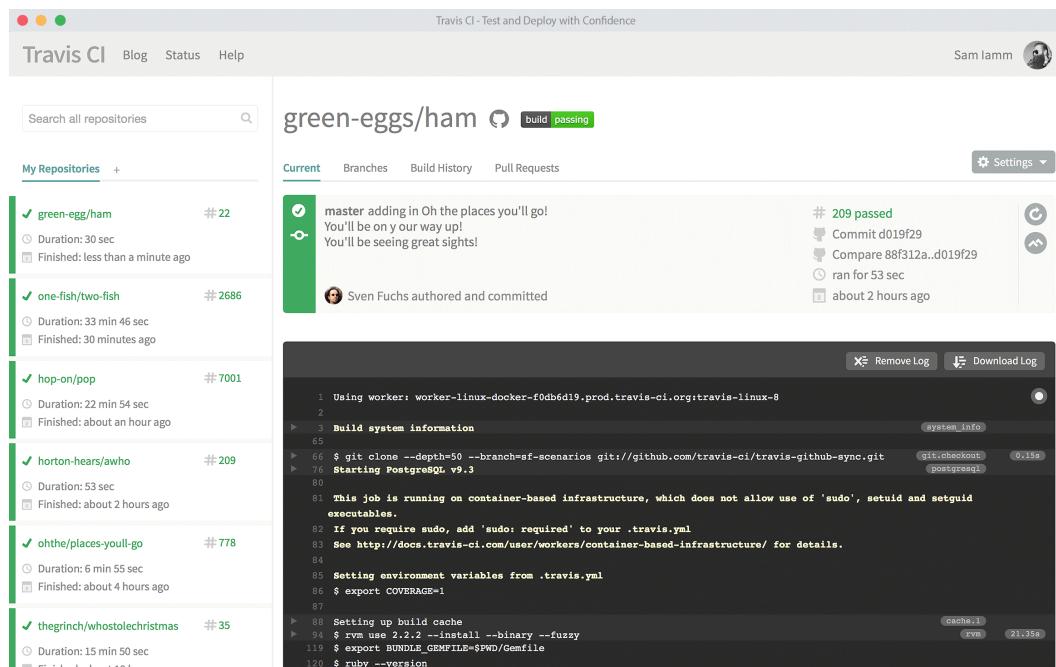


figura 2.2: Pannello di controllo travis

Per lo sviluppo del codice l'*IDE* scelto è stato IntelliJ in quanto:

- Già precedentemente utilizzato;
- Ambiente di sviluppo compatibile con i maggiori linguaggi *JVM-based*;
- Integrazione con Git;
- Integrazione per lo sviluppo in Scala con Sbt.

Per la gestione progetti ho utilizzato il *build tool* Sbt, perché nell'ambiente di sviluppo di Scala è la scelta predominante ed è compatibile con il *framework* utilizzato *Play Framework*. Tra le sue caratteristiche si può notare:

- Definizione del file di configurazione in Scala;
- Compatibilità con Scaladoc per documentazione;
- Supporto per progetti misti Java/Scala;
- Supporto per i maggiori *tool* di *testing*;
- Supporto per la gestione delle librerie.

Per la realizzazione dei diagrammi *UML* (*Unified Modeling Language*) ho utilizzato Astah Professional², uno strumento già utilizzato in precedenza e perché compatibile con gli ultimi standard *UML 2.0*.

²<http://astah.net/>.

2.3.3 Vincoli tecnologici

Scala

Il principale linguaggio di programmazione adottato è Scala³ (acronimo di *"Scalable Language"*). E' stato studiato per poter interoperare con *Java Runtime Environment (JRE)*, quindi è possibile scrivere applicazioni sia in Java e Scala che eseguono nella stessa *Java Virtual Machine (JVM)*. E' un linguaggio multi paradigma, che integra concetti provenienti dai linguaggi della programmazione ad oggetti e della programmazione funzionale. Scala è un linguaggio completamente *object oriented*, infatti ogni elemento del linguaggio è un oggetto, compresi i tipi primitivi e le funzioni. Inoltre è un linguaggio funzionale, include funzioni di prima classe e una libreria con strutture dati immutabili efficienti.

Play Framework

Play è un *framework open source*, scritto in Java e Scala, che permette l'implementazione di *web application* con il *pattern model-view-controller*. Questo *framework* trae ispirazione ed è simile a Ruby on Rails e Django.

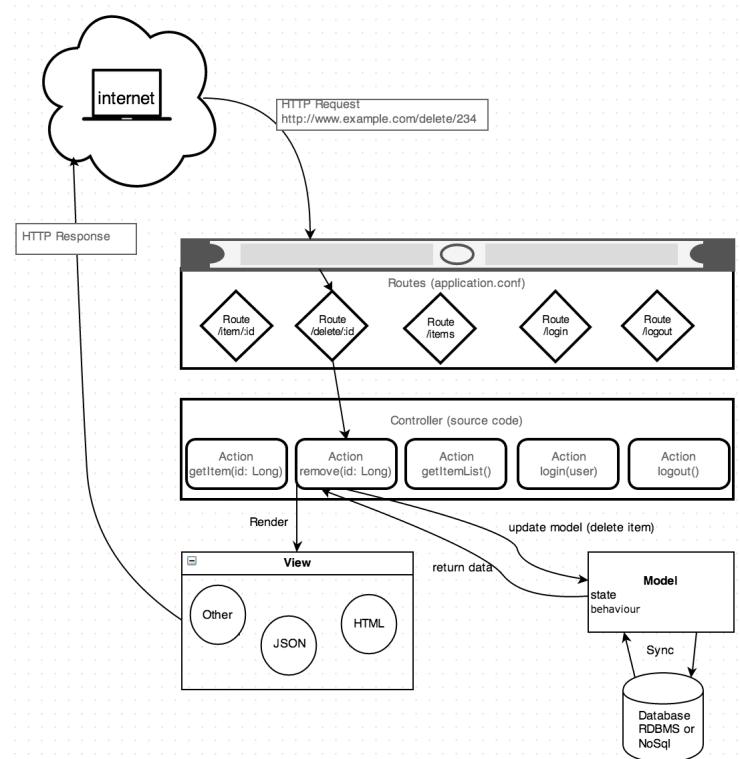


figura 2.3: Tipica architettura di un'applicazione che utilizza Play Framework. Immagine tratta da [Omer Haderi Blog](#).

³<http://www.scala-lang.org/>.

Le peculiarità che lo contraddistinguono dagli altri *framework* sono le seguenti:

- Privo di stato: non viene mantenuta alcuna sessione sul server relativa ai dati dell'utente corrente;
- Metodi statici: tutti i metodi dei *controller* che vengono invocati dal *framework* sono statici, o nel caso in cui si usi la versione Scala, sono funzioni di oggetti di Scala;
- Gestione asincrona dell'*input* e dell'*output*: grazie all'uso di Netty Server, Play può gestire le richieste lunghe in modo asincrono;
- Architettura modulare: come per Rails e Django, ci sono i moduli;
- Supporto nativo per Scala: non solo Play è fatto internamente in Scala, ma espone anche delle interfacce Scala. Le interfacce Java sono state messe appositamente in *package* diversi affinché possano seguire le convenzioni di Java.

Al suo interno permette una implementazione delle *views*, specificando l'interfaccia con un linguaggio di *markup* (e.g. Html, Xml, Json) in combinazione con del codice Scala. Vengono fornite molte classi di *utility* come ad esempio Play Json Library che permette di operare con dei dati in formato Json, oppure un adattamento della libreria Specs2 che permette un supporto per il *testing* di una applicazione in Play. Una tipica applicazione in Play funziona convogliando tutte le *HTTP Request* nella componente *Router* che invocherà a sua volta il *Controller* pre-selezionato. Il *Controller* interagisce con il *Model* che a sua volta restituisce un risultato. In genere questo risultato verrà poi *renderizzato* con la *View* e fatto ritornare tramite un *HTTP Response*.

OrientDB

Per la persistenza dei dati ho utilizzato OrientDb⁴, un *Database Management System* (DBMS) multi-modello con una licenza *open source*. Supporta i seguenti modelli di persistenza:

- *Graph*: modello a grafo;
- *Document*: modello a documenti come MongoDB;
- *Object*: modello ad oggetti che permette di memorizzare direttamente i *POJO*.

Inoltre supporta modelli di definizione dello schema come *schema-less*, *schema-full* e *schema-mixed*. Per motivi prestazionali non permette le *join*, le relazioni sono gestite come in un database a grafo, ovvero delle connessioni dirette tra i *record*. Questa caratteristica lo rende adatto nel contesto dei *Big Data*, infatti la complessità computazionale per gestire una creazione di un *link* è pari a O(1). Tra le caratteristiche più importanti di OrientDb:

- Scalabile: ovvero è possibile aggiungere al sistema un altro *server* per aumentarne le prestazioni;
- Transazionale: supporta transazioni completamente *ACID* garantendo che tutte le transazioni vengano processate in modo affidabile e nel caso di *crash* viene ripristinato lo stato precedente(*rollback*);
- Polyglot: supporta *query* in un dialetto *SQL-like*, oppure delle *API REST* che forniscono l'accesso ai dati.

⁴<http://orientdb.com/>.

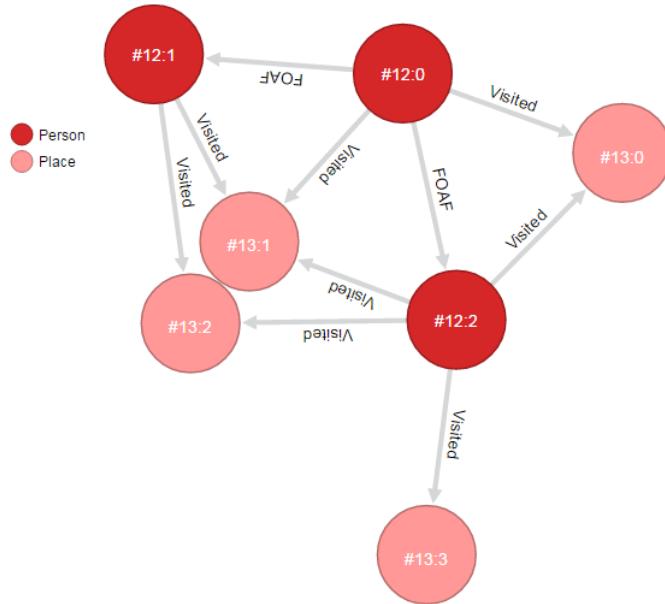


figura 2.4: Esempio di modellazione in OrientDb. Immagine tratta da [Scalac Blog](#)

2.4 Obiettivi personali

Questa sezione elenca gli obiettivi personali che hanno portato allo svolgimento dello *stage*.

Ho conosciuto l’azienda durante l’evento *Stage-IT*. Questo evento mi ha dato la possibilità di approfondire molte proposte di *stage*, da parte aziende provenienti da tutta la provincia di Padova e non. In seguito a tutte le proposte di *stage* disponibili ho accettato la proposta di Nextep S.r.l. per i seguenti motivi:

- Il progetto permetteva di lavorare con tecnologie innovative che mi avrebbero fornito un valore professionale aggiunto;
- Il progetto proposto era interessante perché toccava rami dell’intelligenza artificiale e dei sistemi di raccomandazione;
- L’utilizzo di Scala, in modo da poter padroneggiare un linguaggio con un paradigma funzionale in forte ascesa di popolarità.

Capitolo 3

Il progetto di stage

Questo capitolo tratta dettagliatamente delle attività di pianificazione, studio, analisi, progettazione e sviluppo, svolte durante lo stage.

3.1 Organizzazione

Per conseguire tutti gli obiettivi fissati, insieme al *tutor* aziendale ho stimato un ammontare di 300 ore. Ho pianificato le attività a monte dello stage in modo da distribuire il carico lavorativo in 37.5 ore settimanali, ovvero 7.5 ore giornaliere. In seguito ho rimodulato il piano per focalizzarmi nello sviluppo del modulo Tres.

3.1.1 Pianificazione

Lo stage è cominciato il giorno 5/10/2015 ed è terminato il 27/11/2015. Le attività previste sono state suddivise con cadenza settimanale con questo ordine:

- **Prima settimana:**
 - Studio del *database* OrientDB;
 - Studio del linguaggio di programmazione Scala;
 - Studio del *framework* Play;
 - Studio del progetto preesistente;
- **Seconda e terza settimana:** migrazione dalla tecnologia MongoDB a OrientDB e *porting* all'ultima versione di Play;
- **Quarta settimana:** analisi dei requisiti del nuovo modulo;
- **Quinta settimana:** progettazione architettonale;
- **Sesta e settima settimana:** implementazione modulo;
- **Ottava settimana:** test.

3.2 Analisi dei requisiti

3.2.1 Requisiti

Per l'analisi dei requisiti ho svolto molteplici incontri insieme al *tutor* aziendale per delineare quali funzionalità dovesse offrire il sistema. Il *tutor* mi ha fornito molti esempi ed illustrazioni per capire al meglio quali requisiti il sistema dovesse soddisfare. Una volta definito le funzionalità ho trascritto i requisiti in un file testuale, in un formato organizzato tabellare. Il file è stato poi sottoposto a versionamento all'interno del *repository* per permettere operazioni di consultazione e modifica all'evolversi dei requisiti. La notazione scelta per suddividere i requisiti è la seguente:

$$R[Importanza][Tipologia][Codice]$$

- Importanza può assumere i seguenti valori:
 - **OBB:** requisito obbligatorio. Il soddisfacimento è necessario per il raggiungimento degli obiettivi dello stage;
 - **DES:** requisito desiderabile. L'implementazione non è fondamentale, ma dà valore aggiunto al prodotto.
- Tipologia può assumere i seguenti valori:
 - **F:** requisiti funzionali. Specifica una funzionalità che il software deve avere;
 - **V:** requisiti di vincolo. Specifica il vincolo che il software deve avere;
 - **P:** requisiti di prestazione. Specifica un vincolo di performance che il software deve fornire.
- Codice è un id numerico univoco.

Tabella requisiti:

Codice	Descrizione
ROBBF1	Il sistema deve permettere l'inserimento di un nuovo comportamento
ROBBF2	Il sistema deve restituire un messaggio di segnalazione in caso di errore durante il salvataggio
ROBBF3	I dati ricevuti devono essere in formato Json
ROBBF4	I dati inviati devono essere in formato Json
ROBBF5	Il sistema deve restituire un messaggio di errore in caso di Json non valido
ROBBF6	Devono essere inseriti almeno 100 comportamenti per produrre la raccomandazione

ROBBF7	Il sistema deve restituire un messaggio di conferma quando l'algoritmo è pronto a produrre una raccomandazione
ROBBF8	Il sistema deve permettere il calcolo dell'entropia del dataset
ROBBF9	Il sistema deve permettere il calcolo del guadagno di informazione del dataset
ROBBV10	Il sistema deve permettere la costruzione ricorsivamente dell'albero di decisione
ROBBV11	Il sistema deve permettere la creazione di tanti nodi quanti sono i possibili valori dell'attributo scelto
ROBBV12	L'algoritmo di costruzione dell'albero deve fermarsi quando tutte le istanze di un nodo appartengono alla stessa classe
ROBBF13	Il sistema deve permettere di ricevere una richiesta per classificare un comportamento vuoto
ROBBF14	Il sistema, per la richiesta con un comportamento vuoto, deve restituire tutte le raccomandazioni con le percentuali calcolate in base al dataset
ROBBF15	Il sistema deve permettere di classificare un comportamento dell'utente
ROBBF16	Il sistema deve fornire una raccomandazione o più in base alla classificazione del comportamento
ROBBF17	Il sistema deve fornire insieme alla raccomandazione la percentuale di adeguatezza
ROBBV18	Il sistema deve permettere la modifica dell'albero di decisione ogni 24 ore
ROBBV19	Devono essere rispettate le metriche sulla stesura del codice riportate nella sezione 3.6???
RDESC20	Devono essere implementati nuovi algoritmi basati sulla teoria dei giochi

RDESV21	Devono essere implementati nuovi algoritmi di clustering
RDESV22	Deve essere sviluppato il pannello di amministrazione
ROBBP23	Il database deve essere in grado di memorizzare almeno 100.000 record al secondo

tabella 3.1: Tabella dei requisiti

3.2.2 Casi d'uso

In questa sezione riporto i casi d'uso principali. Tres è concepito per operare con Bdrim, una piattaforma di raccolta dati sviluppato in collaborazione con Allos. L'unico attore coinvolto nell'utilizzo di Tres è Bdrim che espone un caso d'uso principale. Per ogni caso d'uso ho specificato gli attori coinvolti, pre e post condizioni, descrizione, scenario principale e se previsto lo scenario alternativo.

UC0

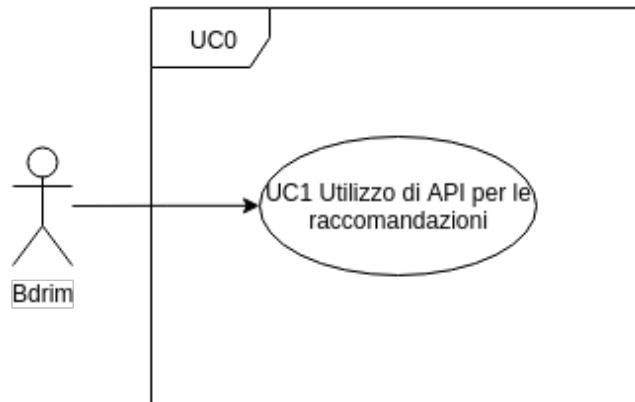


figura 3.1: UC0: Tres

- **Attori:** Bdrim
- **Descrizione:** Bdrim usufruisce delle API per ottenere una raccomandazione.
- **Precondizione:** Il server è stato avviato.
- **Postcondizione:** Il server risponde alle richieste HTTP effettuata dall'utente.

UC1 Revisionare!!!!

- **Attori:** Bdrim

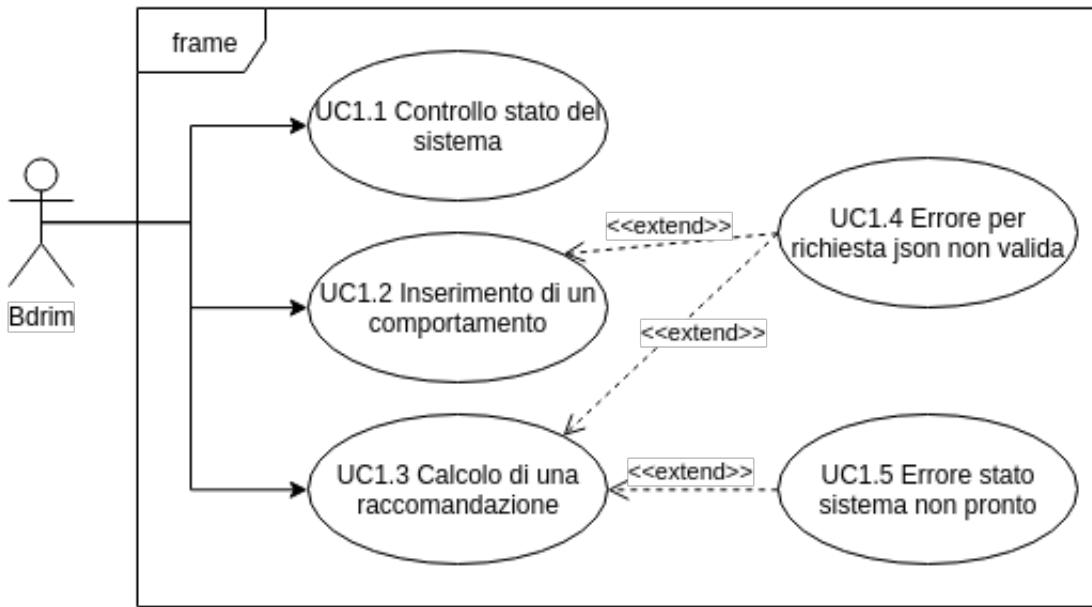


figura 3.2: UC1: Tres

- **Descrizione:** Questo caso d'uso illustra le funzionalità offerte dalle *API* di raccomandazione a Bdrim.
- **Precondizione:** Tres è stato attivato.
- **Flusso principale:**
 - 1 Bdrim può richiedere lo stato del sistema;
 - 2 Bdrim può inserire un comportamento;
 - 3 Bdrim può richiedere la raccomandazione.
- **Flusso alternativo:**
- **Postcondizione:** il sistema risponde con una risposta HTTP alla richiesta, ritornando i dati in formato Json.

3.3 Progettazione architetturale

Questa sezione descrive l'attività di progettazione. Illustra l'architettura generale del sistema, le relazioni interne, la struttura del database e infine i *design pattern* utilizzati.

3.3.1 Visione ad alto livello

Il contesto di utilizzo dell'applicativo è la piattaforma di Bdrim. Questa piattaforma permette di gestire nuovi moduli per introdurre nuove funzionalità. Ogni modulo deve essere assolutamente isolato per poter essere facilmente modificabile e interscambiabile con altri moduli che migliorino le sue funzionalità. Quindi la visione generale del sistema si riconduce ad un modello *client-server*.

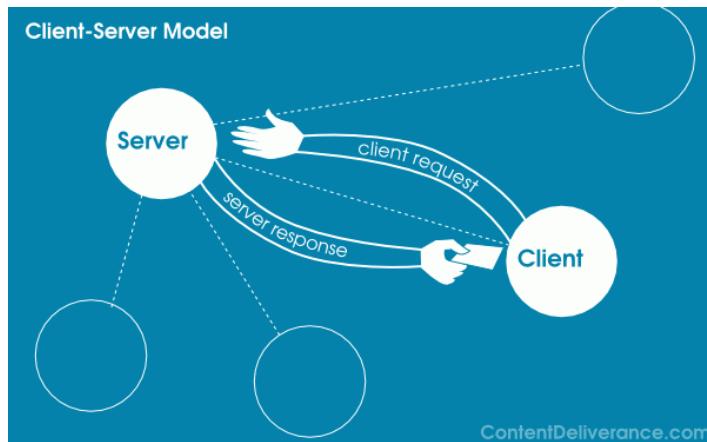


figura 3.3: Visione generale architettura *client-server*. Immagine tratta da [Content Deliverance](#)

3.4 Progettazione di dettaglio

3.5 Verifica e Validazione

3.5.1 Resoconto risultati

Metriche misurate

Qui di seguito sono riportate le metriche software misurate:

- **Complessità ciclomatica:** per il calcolo della complessità ciclomatica ho utilizzato Sbt, al suo interno è disponibile il *tool* styleCheck. Questo strumento segnala con un *warning* se la complessità supera il valore 10. Il *tool* ha restituito esito positivo e non ha segnalato alcun *warning*
- **Attributi per classe:** ho rilevato per questa metrica come valore medio 2, mentre come valore massimo ho rilevato 4.
- **Copertura commenti:** il valore di copertura è pari al 100%
- **Copertura dei test:** il valore di copertura dei test è del 73%

Test di unità

Qui di seguito sono riportati tutti i test di unità eseguiti e il loro relativo esito:

Test	Descrizione	Stato
TU1	Viene verificato che il sistema salvi correttamente un comportamento all'interno del database	Esito positivo
TU2	Viene verificato il corretto caricamento di una lista di <i>widgetTag</i> distinti	Esito positivo

TU3	Viene verificato il corretto caricamento di una lista di <i>item</i> distinti	Esito positivo
TU4	Viene verificato il corretto caricamento di una lista di <i>behavior</i> con una <i>interaction</i> specifica	Esito positivo
TU5	Viene verificato il corretto caricamento di una lista di <i>behavior</i> con <i>tag</i> e <i>action</i> specifici	Esito positivo
TU6	Viene verificato il corretto caricamento di una lista di <i>action</i> distinti	Esito positivo
TU7	Viene verificato il corretto calcolo dell'entropia	Esito positivo
TU8	Viene verificato il corretto calcolo del guadagno di informazione	Esito positivo
TU9	Viene verificato che venga creata la raccomandazione corretta da fornire	Esito positivo
TU10	Viene verificato il corretto calcolo delle probabilità	Esito positivo

tabella 3.2: Tabella dei test di unità

Test di integrazione

Qui di seguito sono riportati tutti i test di integrazione eseguiti e il loro relativo esito:

Test	Descrizione	Componente	Stato
TI1	Viene verificato che il sistema crei correttamente un albero di decisione	algorithm	Esito positivo
TI2	Viene verificato che il sistema aggiorni l'albero di decisione ogni 24 ore	algorithm	Esito positivo
TI3	Viene verificata la corretta creazione del <i>training set</i>	algorithm	Esito positivo

tabella 3.3: Tabella dei test di integrazione

Test di sistema

Ho effettuato alcuni test di sistema inerenti alle funzionalità di inserimento dati, calcolo di una raccomandazione e infine il calcolo probabilità.

- **Inserimento dati:** viene verificato il corretto comportamento del sistema alla richiesta di inserimento dati;

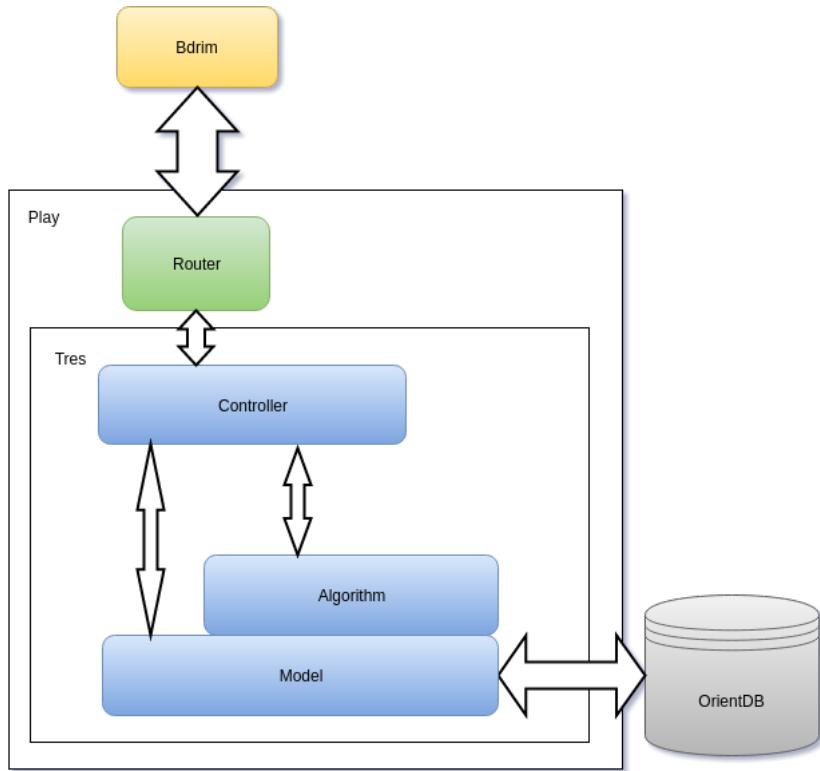


figura 3.4: Visione generale dell'architettura

- **Calcolo raccomandazioni:** viene verificato il calcolo della raccomandazione da fornire;
- **Calcolo probabilità:** viene verificato il calcolo delle probabilità associate ai vari *item*.

Capitolo 4

Valutazione Retrospettiva

Questo capitolo riporta un bilancio finale su quanto svolto durante lo stage.

4.1 Bilancio sui risultati

In questa sezione viene fatto un riassunto degli obiettivi aziendali e gli obiettivi personali raggiunti.

4.1.1 Obiettivi conseguiti

4.1.2 Obiettivi personali

4.2 Bilancio formativo

In questa sezione viene riportato le competenze acquisite durante lo stage.

4.3 Distanza tra formazione universitaria e lavoro

In questa sezione viene esposto la valutazione personale della distanza tra la formazione ricevuta durante il corso di studi e lo stage formativo.

Appendice A

Appendice A

Citazione

Autore della citazione

Bibliografia

Riferimenti bibliografici

Siti Web consultati