

Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

CORSO DI LAUREA IN INFORMATICA



**Analisi di motori di ricerca Open Source
per siti web informativi**

Tesi di laurea triennale

Relatore

Prof. Tullio Vardanega

Laureando

Daniel De Gaspari

ANNO ACCADEMICO 2016/2017

placeholder con citazione.

— Oscar Wilde

Sommario

Il presente documento descrive il lavoro svolto durante il periodo di stage dal laureando Daniel De Gaspari, della durata di circa trecento ore, presso l'azienda InfoCamere S.C.p.A. di Padova (PD).

Gli obiettivi da raggiungere erano molteplici.

Lo scopo dello stage consisteva nell'analisi delle caratteristiche dei motori di ricerca [open source](#) nell'ambito dei siti web di tipo informativo.

In primo luogo era richiesto un approfondimento delle caratteristiche istituzionali dei siti web delle Camere di Commercio.

Successivamente, l'azienda richiedeva di analizzare le potenzialità e specificità di motori di ricerca [Sori](#) e [ElasticSearch](#).

Il passo successivo consisteva nel realizzare un prototipo di un sito web in tecnologia [Drupal](#), con i due motori di ricerca precedentemente citati.

Infine, era richiesta una relazione finale delle potenzialità emerse nell'utilizzo dei due motori di ricerca.

I primi due capitoli del presente documento hanno lo scopo di presentare il contesto aziendale in cui è stato sostenuto lo stage e di spiegare come il progetto di stage si renda utile all'interno della strategia aziendale. Il terzo capitolo documenta invece lo svolgimento dello stage descrivendo le attività che sono state portate a termine, i punti salienti del progetto stesso e le principali scelte attuate. Il quarto ed ultimo capitolo presenta infine una valutazione dello svolgimento dello stage rispetto agli obiettivi aziendali e alle conoscenze acquisite dallo studente.

"Citazione"

— Confucius

Ringraziamenti

Ringraziamenti

Padova, Dicembre 2017

Daniel De Gaspari

Indice

1	L'azienda	1
1.1	Il Profilo Aziendale	1
1.1.1	Le origini: Cerved	1
1.1.2	Anni '90: InfoCamere	2
1.1.3	Servizi offerti	3
1.2	Organizzazione aziendale	4
1.3	Processi aziendali	5
1.3.1	La fornitura	5
1.3.2	Ciclo di vita dei siti web	6
1.3.3	Auditing	6
1.3.4	Manutenzione	6
1.4	Tecnologie utilizzate	7
1.4.1	Ambiente di sviluppo	7
1.4.2	Gestione di versione	9
1.4.3	Comunicazioni	9
1.5	Rapporto con l'innovazione	9
2	Lo stage	11
2.1	Gli stage in azienda	11
2.2	L'offerta di stage	12
2.2.1	Presentazione del progetto	12
2.2.2	Obiettivi posti dall'azienda	13
2.2.3	Vincoli	15
2.3	Obiettivi personali	16
3	Resoconto dello stage	19
3.1	Individuazione dei motori di ricerca	19
3.2	Pianificazione	19
3.3	I siti istituzionali delle Camere di Commercio	19
3.3.1	Funzionalità di ricerca attuali	19
3.3.2	Possibile evoluzione	19
3.4	Ricerca nativa Drupal	19
3.4.1	Introduzione a Drupal	19
3.4.2	Ricerca di base e avanzata	19
3.4.3	Ricerca con Search API	20
3.4.4	Considerazioni di Drupal nativo	20
3.5	Ricerca con Solr	20
3.5.1	Introduzione a Solr	20

3.5.2	Principali funzionalità di ricerca	20
3.5.3	Integrazione con Drupal	20
3.6	Ricerca con ElasticSearch	20
3.6.1	Introduzione a ElasticSearch	20
3.6.2	Principali funzionalità di ricerca	20
3.6.3	Integrazione con Drupal	20
3.7	Considerazioni finali sui motori di ricerca esaminati	21
4	Valutazione retrospettiva	23
4.1	Bilancio degli obiettivi	23
4.1.1	Aziendali	23
4.1.2	Personali	23
4.2	Conoscenze acquisite	23
4.3	Mondo del lavoro e università a confronto	23
	Glossario	25
	Acronimi	27
	Bibliografia	29

Elenco delle figure

1.1	Storyline InfoCamere	1
1.2	Logo infocamere	2
1.3	<i>Esempio di attuale utilizzo del registro imprese.</i> URL: http://www.registroimprese.it/	2
1.4	<i>Dislocazione delle Camere di Commercio Italiane e relative sedi distaccate.</i> URL: https://www.infocamere.it/noi-e-le-camere-di-commercio	3
1.5	<i>Servizi telematici offerti dal SUAP di Venezia.</i> URL: http://www.impresainungiorno.gov.it/web/guest/comune?codCatastale=L736	4
1.6	Organigramma aziendale	5
1.7	Ciclo di vita dei siti web	6
1.8	Verbale firmato da entrambi i referenti tecnici	6
1.9	<i>Nuovo istanza di un sito Drupal.</i> URL: https://www.drupal.org/documentation/install/beginners	8
1.10	Pannello di controllo di Acquia Dev Desktop 2	8
1.11	Client di posta elettronica Zimbra	9
2.1	Esempi di funzionalità di ricerca offerte dal sito di Amazon	12
2.2	Milestone e metodologia di lavoro	14
2.3	Tipologie di vincoli del progetto	15

Elenco delle tabelle

2.1	Obiettivi dello stage	15
-----	---------------------------------	----

Capitolo 1

L'azienda

1.1 Il Profilo Aziendale

1.1.1 Le origini: Cerved

Nata inizialmente come Cerved (Centro Regionale Veneto Elaborazione Dati), InfoCamere S.C.p.A. è stata fondata nel Dicembre del 1974 a Padova dal Professor Mario Volpato, allora Presidente della Camera di Commercio di Padova e Professore di Calcolo delle probabilità all'Università di Padova.

L'obiettivo era di raccogliere e conservare i dati ufficiali anagrafici e amministrativi delle imprese della provincia di Padova in un modo nuovo rispetto a quanto previsto fino ad allora: la conservazione di quei dati su un registro cartaceo, come si faceva dal medioevo ai tempi delle comunità dei mercanti, non bastava più a garantire l'efficienza del mercato e a stimolare lo sviluppo economico.

Le prime tecnologie informatiche aprivano nuovi orizzonti al trattamento massivo e veloce dei dati. Evolveva rapidamente la concezione di una gestione intelligente delle notizie amministrative sulla vita delle imprese, per trasformarle in informazioni rielaborabili ed utilizzabili in modi nuovi da tutti. Nasceva l'idea di valorizzare i dati ufficiali forniti dalle imprese, restituendoli al mercato e alle imprese stesse come informazioni utili per accrescere la propria competitività e progettare lo sviluppo.

Si gettava il seme dell'efficienza nell'organizzazione delle Camere di Commercio, una base nuova per costruire un patto trasparente e vantaggioso tra imprese e Pubblica Amministrazione.

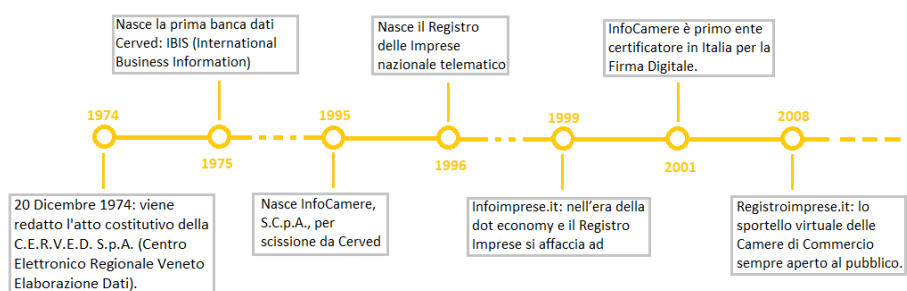


Figura 1.1: Storyline InfoCamere

1.1.2 Anni '90: InfoCamere

All'inizio degli anni '90 aumentava sempre più la competizione globale e le sfide per portare l'Italia nella modernità. A tal fine, nel 1993, venne emanata una riforma (legge 29 Dicembre 1993, n. 580) che attribuiva alle Camere un'autonomia rispetto al governo centrale, mediante attribuzione della potestà statutaria e di autonomia finanziaria, oltre al riconoscimento del ruolo finalizzato alla pubblicizzazione delle imprese. Le Camere di Commercio Italiane hanno così modo di vedere un profondo rinnovamento in vari ambiti e in particolar modo nell'ambito tecnologico.



Figura 1.2: Logo infocamere

Nel 1995, per scissione da Cerved, nasce InfoCamere che raccoglie la sfida di realizzare il Registro delle imprese. Previsto dal codice civile fin dal 1942 e mai attuato, in due anni, con uno di anticipo sulle previsioni, il risultato è raggiunto: prende vita il primo esempio in Europa di registro pubblico sulle imprese totalmente telematico; assieme ad un ecosistema di servizi sviluppati attorno al Registro delle imprese, è stato possibile semplificare i processi tra le imprese stesse e la Pubblica Amministrazione.

AUTOENERGY SRL

Nome Impresa: Autoenergy Srl
 Indirizzo Impresa: Bassano Del Grappa (VI) Largo Parolini 52 [Mostra mappa](#)
 Forma giuridica (generico): Società di capitale
 Attività: Officina di produzione di energia elettrica mediante un gruppo cogeneratore alimentato ad olio vegetale
 Registrata: ☒

Per visualizzare la PEC inserisci il codice nel box e premi Mostra

PEC: [MOSTRA](#)

PER INFORMAZIONI ANAGRAFICHE COMPLETE [ACCEDI](#) o [REGISTRATI](#)
 Numero REA, Codice Fiscale, Forma Giuridica Completa, Descrizione Estesa dell'attività

ACQUISTA I PROSPETTI UFFICIALI DAL REGISTRO IMPRESE

VISURA ORDINARIA	€ 5,00
VISURA D'IMPRESA IN INGLESE	€ 5,00
VISURA STORICA	€ 6,00
FASCICOLO	€ 10,00
FASCICOLO STORICO	€ 11,00
ULTIMO BILANCIO (2016)	€ 2,50

Figura 1.3: Esempio di attuale utilizzo del registro imprese. URL: <http://www.registroimprese.it/>

A quella sfida ne seguono altre che rispondono ai nomi di 'firma digitale', 'posta elettronica certificata', 'comunicazione unica', ecc... Attraverso InfoCamere, servizi e tecnologie digitali di frontiera diventavano patrimonio quotidiano della comunità delle imprese e dei professionisti, influenzando sulle abitudini

di lavoro di migliaia di italiani e stimolando i processi di innovazione nella Pubblica Amministrazione.

1.1.3 Servizi offerti

InfoCamere S.C.p.A. è la società consortile di informatica delle Camere di Commercio Italiane. Ha realizzato e gestisce il sistema telematico nazionale che collega tra loro tutte le Camere di Commercio, oltre alle rispettive sedi distaccate.



Figura 1.4: *Displocazione delle Camere di Commercio Italiane e relative sedi distaccate.*
URL: <https://www.infocamere.it/noi-e-le-camere-di-commercio>

Sua funzione istituzionale è anche la gestione e divulgazione del patrimonio informativo camerale, con particolare riferimento alle informazioni derivanti dal Registro delle imprese.

Le banche dati camerali sono rese disponibili direttamente a imprese, pubbliche amministrazioni, professionisti e cittadini tramite il portale delle Camere di Commercio.

La Società fornisce alle pubbliche amministrazioni l'accesso al Registro Imprese, assicurando loro l'accessibilità dei dati senza oneri, salvo quelli per la fornitura telematica e i servizi a valore aggiunto.

Tramite il sito del Registro delle imprese si può accedere agli strumenti per lo svolgimento delle pratiche telematiche, tra cui la Comunicazione Unica per l'attività d'impresa, valida anche per Agenzia delle Entrate, INPS, INAIL e Albo Artigiani. Il Registro Imprese è inoltre uno strumento di trasparenza amministrativa che fornisce un contributo importante nella lotta contro la criminalità economica. L'azienda ha infatti sviluppato per le autorità investigative alcuni servizi che, in questa direzione, consentono analisi mirate per monitorare fenomeni anomali.

InfoCamere ha realizzato, per conto delle Camere di Commercio, l'infrastruttura tecnologica che garantisce il corretto funzionamento degli Sportelli Unici per le Attività Produttive (SUAP).

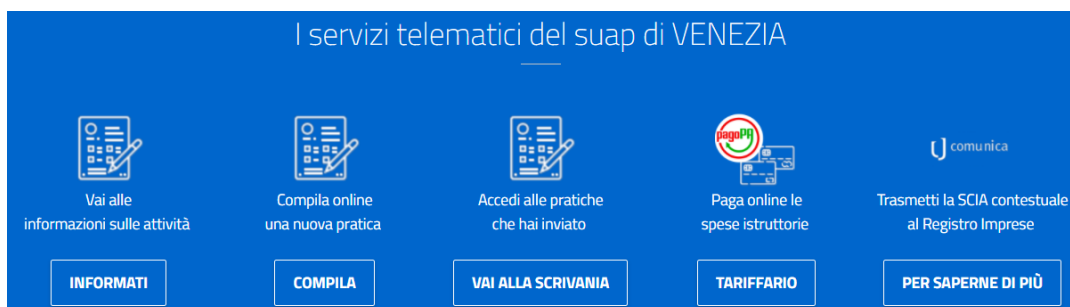


Figura 1.5: Servizi telematici offerti dal SUAP di Venezia. URL: <http://www.impresainungiorno.gov.it/web/guest/comune?codCatastale=L736>

Tra le realizzazioni di InfoCamere per il Sistema camerale vi è anche la procedura informatica che consente di gestire il servizio di conciliazione online (Concilia Camera), fornendo così ad imprese, consumatori e professionisti uno strumento che permette di ricevere assistenza specializzata nel raggiungimento di un accordo per risolvere in modo semplice, rapido, economico e sicuro una controversia, evitando di ricorrere alla giustizia ordinaria.

InfoCamere è inoltre l'Autorità di Certificazione Nazionale che rilascia i certificati digitali delle Carte Tachigrafiche. La società si è dotata di un Sistema di Gestione della Sicurezza delle Informazioni certificato secondo lo standard ISO/IEC 27001, avendo conseguito nel 2012 la prima certificazione di conformità ISO/IEC 27001:2005 e a Marzo 2015 la ricertificazione secondo la nuova versione ISO/IEC 27001: 2013.

1.2 Organizzazione aziendale

In InfoCamere S.C.p.A. è possibile individuare 4 aree direzionali di maggior interesse:

- Servizi alle Camere di Commercio;
- Servizi a imprese, Pubblica Amministrazione, professionisti e altri utenti;
- Tecnologie e impianti;
- Governo progetti, innovazione ed azienda digitale.

Nello specifico, durante lo stage, ho preso parte all'area direzionale "Servizi alle Camere di Commercio". In quest'ultima è possibile individuare:

- Area commerciale: si occupa degli accordi commerciali con le Camere di Commercio;
- Sviluppo ed erogazione servizi alle Camere: attua quanto accordato dall'area commerciale.

In particolare, sono stato assegnato all'unità organizzativa denominata "Camera Digitale", che risponde allo "Sviluppo ed erogazione servizi alle Camere". Questa unità organizzativa si occupa della digitalizzazione delle Camere, sia per quanto riguarda la gestione documentale, rispettando le norme riguardanti la conservazione dei documenti, sia per quanto riguarda i siti web informativi delle Camere di Commercio.

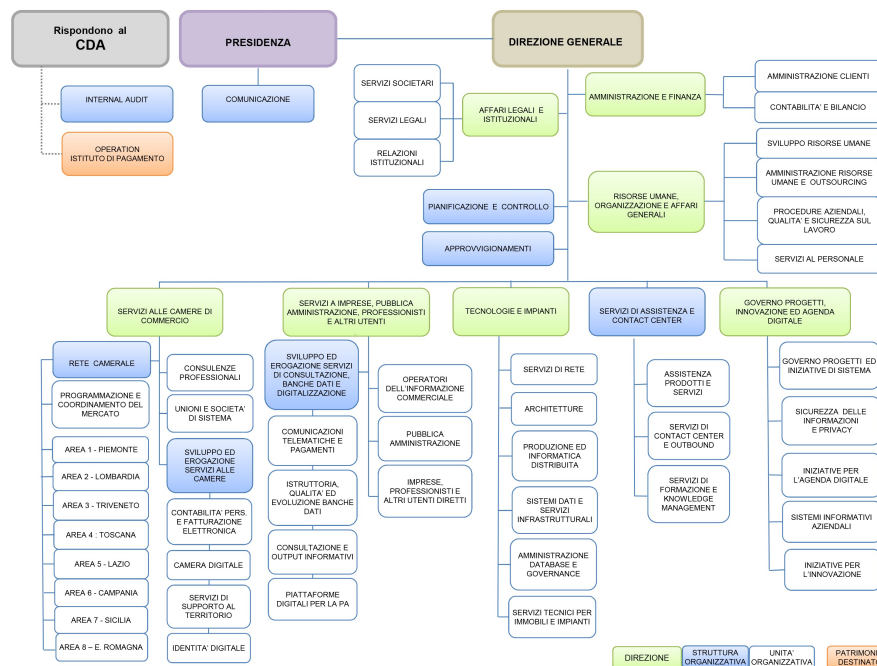


Figura 1.6: Organigramma aziendale

1.3 Processi aziendali

In questa sezione presenterò i processi aziendali che principalmente coinvolgono l'unità organizzativa "Camera Digitale", con la quale ho avuto modo di confrontarmi.

1.3.1 La fornitura

L'azienda mette a disposizione dei clienti due differenti tipologie di prodotto e, nello specifico, di siti web:

- Listino: rappresentano prodotti la cui forma, contenuto e funzione siano idonei alla replicazione;
- Commessa: rappresentano prodotti la cui forma, contenuto e funzione vengono fissate dal cliente.

La contrattazione con il cliente viene interamente gestita dall'area commerciale.

Se la tipologia di prodotto scelta dal cliente è di tipo "Commessa", si opera una raccolta dei requisiti a cui segue una proposta commerciale al cliente.

In entrambe le tipologie di prodotto, quando l'offerta è stata concordata, la Camera di Commercio si occuperà dell'approvazione mediante una delibera pubblica.

Una volta avvenuto l'ingaggio, se il prodotto appartiene alla tipologia "Listino", verrà assegnato al nuovo prodotto un codice identificativo generico; in caso contrario, se la tipologia è "Commessa", al prodotto sarà invece attribuito un codice specifico.

A seguito dell'assegnazione di un nuovo codice identificativo del prodotto da realizzare, è possibile iniziare a sviluppare il sito web.

1.3.2 Ciclo di vita dei siti web

Il ciclo di vita dei siti web prodotti da InfoCamere, è così formato:

- Sviluppo: avviene lo sviluppo dei siti web e delle relative funzionalità; comprende specifici file system e database dedicati allo sviluppo, contenenti dati fittizi, oltre al software che dovrà essere, una volta pronto, distribuito;
- Test: segue lo sviluppo, dal quale prende in input il software prodotto. Il database è lo stesso che era presente in Produzione il giorno precedente a quello considerato;
- Produzione: contiene il software che ha superato i test, e il database è popolato con dati inseriti direttamente dalla Camera di Commercio.

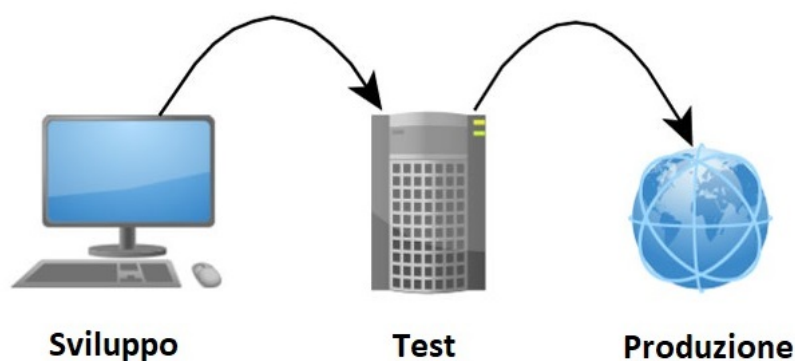


Figura 1.7: Ciclo di vita dei siti web

1.3.3 Auditing

A seguito del caricamento dei contenuti, da parte della Camera, nel sito sviluppato, parteciperanno al collaudo: il commerciale di InfoCamere che ha seguito il lavoro, un referente tecnico di InfoCamere e infine un referente tecnico della Camera di Commercio. Il prodotto di questo processo sarà un verbale del collaudo che dovrà essere firmato e protocollato, oltre a presentare la firma dei due referenti tecnici.

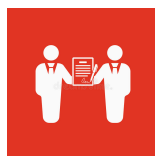


Figura 1.8: Verbale firmato da entrambi i referenti tecnici

1.3.4 Manutenzione

Alla protocollazione del verbale prodotto dal collaudo, si può considerare concluso lo sviluppo del sito e prende avvio l'assistenza al prodotto.

Tipologie di manutenzione

La manutenzione può essere di 3 differenti tipologie:

- Correttiva: vengono corretti bug presenti nel prodotto;
- Adattativa: i siti vengono adeguati a nuove norme in vigore;
- Evolutiva: vengono aggiunte funzionalità al sito; il prodotto subisce un'evoluzione: la natura del prodotto cambia in modo radicale, mantenendo però il prodotto stesso.

Gestione dei ticket

L'assistenza al prodotto viene gestita a più livelli, a cui affluiscono più gruppi di lavoro. Un nuovo ticket, prodotto di una segnalazione di un cliente, può essere gestito in modo autonomo dal primo livello o può essere assegnato ad un gruppo di lavoro appartenente al livello successivo.

Il gruppo di lavoro a cui viene assegnato il nuovo ticket può:

- decidere di rifiutare il ticket: in questo caso, il ticket dovrà essere assegnato ad un altro gruppo di lavoro del secondo livello;
- rimandare ad un altro gruppo di lavoro il ticket, nel caso in cui il gruppo identificato sia ritenuto più adatto a risolvere il problema;
- decidere di utilizzare le proprie competenze per analizzare e cercare di risolvere il problema, fornendo eventualmente una soluzione e un tempo atteso.

I ticket che richiedono una manutenzione evolutiva, comporteranno inoltre uno studio di fattibilità, fatto dal gruppo del secondo livello al quale il ticket è stato assegnato, e un successivo preventivo orario ed economico.

Per i siti web, è garantito ai clienti un pacchetto di ore di supporto specificatamente per i ticket di tipo implementativo. Se il numero di ore preventivato è compreso nelle ore residue del pacchetto garantite al cliente, l'implementazione delle nuove funzionalità può avere luogo. In caso contrario, è necessaria la figura del commerciale per gestire la richiesta di manutenzione evolutiva.

1.4 Tecnologie utilizzate

Di seguito, presenterò le tecnologie con cui sono venuto a contatto durante lo stage.

1.4.1 Ambiente di sviluppo

Drupal

L'unità organizzativa alla quale ho preso parte, "Camera Digitale", realizzava i siti web con [Drupal](#). Si tratta di un [Content Management System](#), rilasciato sotto licenza [open source](#), che permette la creazione di siti Internet, blog e portali, gallerie di immagini, forum di discussione, piattaforme intranet e molto altro. Essa è altresì un'applicazione completamente web based e può quindi essere utilizzata attraverso un semplice browser. E' interamente sviluppato in [PHP](#) e utilizza come base di dati [MySQL](#) in modo nativo.

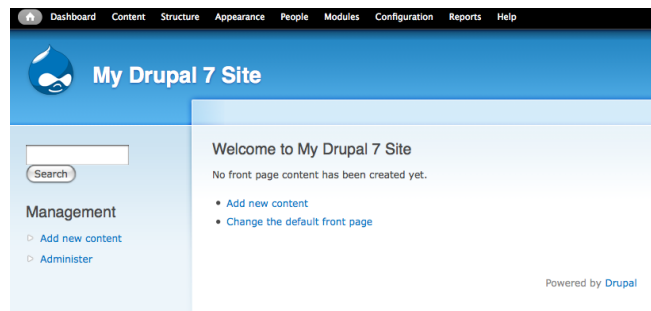


Figura 1.9: Nuova istanza di un sito Drupal. URL: <https://www.drupal.org/documentation/install/beginners>

Acquia Dev Desktop 2

E' un software che permette di realizzare e gestire siti dinamici, che possono accrescere e mutare il proprio contenuto continuamente. Di seguito, sono elencati i più importanti chiave di questa tecnologia:

- Gratuito per uso personale;
- E' il modo più veloce per disporre di un sistema [Drupal](#), fornendo un [DAMP](#) stack installer in modo tale da consentire l'installazione di tutte le componenti necessarie per avviare [Drupal](#), comprendente [Apache](#), [MySQL](#) e [PHP](#);
- Permette la gestione di più istanze [Drupal](#) in modo semplice e veloce. Questa caratteristica risulta essere particolarmente significativa, in quanto ad ogni diversa tecnologia utilizzata per la ricerca è possibile associare una nuova istanza [Drupal](#), così da evitare interferenze tra le tecnologie e mantenere il lavoro ben distinto.

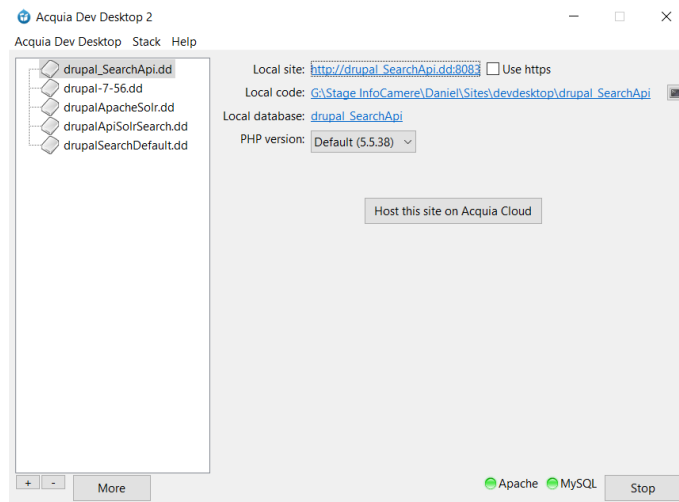


Figura 1.10: Pannello di controllo di Acquia Dev Desktop 2

1.4.2 Gestione di versione

Git è un sistema di controllo di versione distribuito e **open source** ed è lo strumento utilizzato dall'azienda per il versionamento del codice. In particolare, viene utilizzato **Git Extensions**, che rappresenta un'interfaccia grafica per **Git**, permettendone l'utilizzo senza dover ricorrere alla riga di comando. Il materiale prodotto durante lo stage, comprensivo di copie delle istanze **Drupal** create, è stato versionato. In questo modo è possibile prevenire eventuali perdite di dati derivanti dal malfunzionamento della macchina utilizzata, oltre a mantenere le varie versioni di quanto prodotto; in qualunque momento è dunque possibile ricreare l'ambiente di sviluppo di qualsiasi versione e recuperare tutti i prodotti realizzati.

1.4.3 Comunicazioni

Zimbra è il client di posta elettronica utilizzato da InfoCamere S.C.p.A. . Questa tecnologia è **open source** e prevede un sistema di chiamate vocali e messaggistica istantanea. La suite di comunicazione può essere configurata in base alle esigenze personali di ciascuna azienda. Con un servizio cloud installato su server dedicati, tutti gli strumenti di comunicazione aziendale sono facilmente gestibili e ottimizzabili. E' molto utile in ambito aziendale, per il coordinamento e la collaborazione tra colleghi.

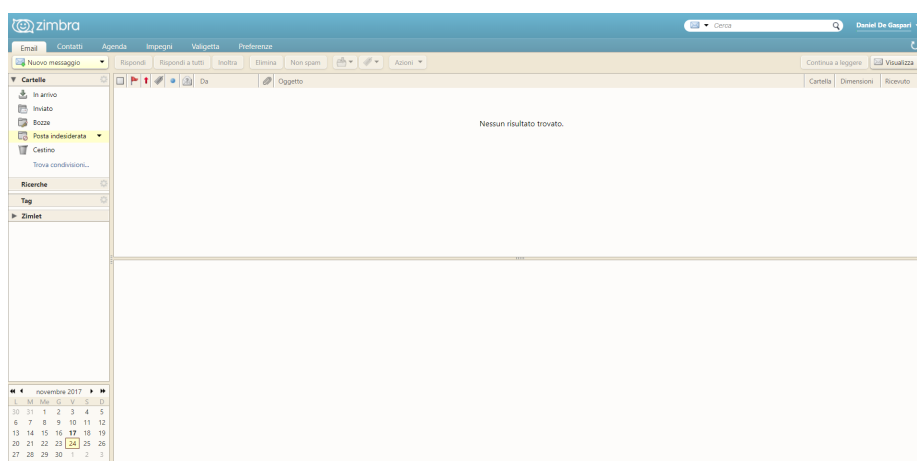


Figura 1.11: Client di posta elettronica Zimbra

1.5 Rapporto con l'innovazione

InfoCamere S.C.p.A. guarda all'innovazione con grande interesse, tanto da avere un'area direzionale che tratta la materia dell'innovazione, come si può vedere nella sezione dedicata all'**organizzazione aziendale**.

Le Camere di Commercio demandano ad InfoCamere il compito di portare innovazione nei servizi che le stesse mettono a disposizione delle imprese, soprattutto nel loro rapporto con la Pubblica Amministrazione, con l'obiettivo di semplificare e alleggerire i costi burocratici e gestionali. In linea con questo principio, InfoCamere persegue la strada della dematerializzazione, della firma digitale, della conservazione sostitutiva

della carta, della fatturazione elettronica e di tutti gli strumenti che la tecnologia e l'informatica offrono per garantire una governance delle procedure amministrative al passo con le esigenze delle imprese e coi tempi.

In quest'ottica rientrano tutti i servizi amministrativi che InfoCamere ha progettato nel tempo per conto delle Camere di Commercio Italiane: dall'invio online delle pratiche, allo Sportello Unico per le Attività Produttive, agli strumenti per le imprese dedicati alla giustizia civile, alla fatturazione elettronica, ai pagamenti verso le Pubbliche Amministrazioni fino a quelli di certificazione digitale, elemento essenziale per il funzionamento dei servizi online.

Capitolo 2

Lo stage

2.1 Gli stage in azienda

Lo stage identifica un periodo lavorativo all'interno di un'azienda, svolto da una persona come se essa fosse un dipendente. L'obiettivo primario dello stage è quello di far apprendere alla persona le competenze necessarie per poter essere poi, eventualmente, assunto, divenendo così un dipendente a tutti gli effetti.

InfoCamere S.C.p.A. si adegua pienamente a questa definizione, reputando gli stage di grande interesse, con il fine di portare innovazione all'interno dell'azienda e possibilmente nuovi dipendenti.

Le motivazioni che spingono dunque l'azienda ad ospitare gli stage, sono molteplici:

- Una prima motivazione è legata alla continua ricerca di innovazione da portare all'interno dell'azienda. Capita spesso che i dipendenti di InfoCamere S.C.p.A. siano impegnati in parallelo in vari progetti, con scadenze anche stringenti, non avendo tempo dunque da dedicare all'approfondimento di tecnologie che potrebbero portare miglioramenti all'azienda e ai suoi prodotti. Una risorsa come uno studente universitario risulta dunque essere di gran interesse, permettendo l'esplorazione di nuove tecnologie, senza però dover rallentare altre attività e progetti in opera;
- Menti derivanti dal mondo universitario hanno le caratteristiche adatte a portare nuovi modi di pensare e di vedere le cose, avendo molte volte la capacità di affrontare i problemi sotto diversi punti di vista rispetto a quelli usualmente trattati aziendalmente. Tutto ciò, assieme all'esperienza che l'azienda può fornire, può evolvere in nuove idee e iniziative per portare un valore aggiunto all'azienda stessa;
- Uno stage che porta ad un soddisfacimento sia dello studente, sia dell'azienda, può far sì che il rapporto tra le due parti continui anche al termine dello stage. In questo modo viene garantito all'azienda un periodo di prova nel quale può decidere se proporre o meno, al termine dello stage, un contratto di assunzione, in modo tale da incrementare o ringiovanire l'organico aziendale;
- Dal punto di vista aziendale, lo stage è conveniente anche sotto l'aspetto economico, in quanto la stessa non è tenuta al pagamento di uno stipendio allo stagista, ed è a sua discrezione l'assegnazione di un rimborso spese. Uno stage

curricolare, come quello da me svolto, produce ulteriori vantaggi economici per l'azienda, data la copertura assicurativa antinfortunistica sul luogo del lavoro, che risulta essere a completo carico dell'Università.

2.2 L'offerta di stage

2.2.1 Presentazione del progetto

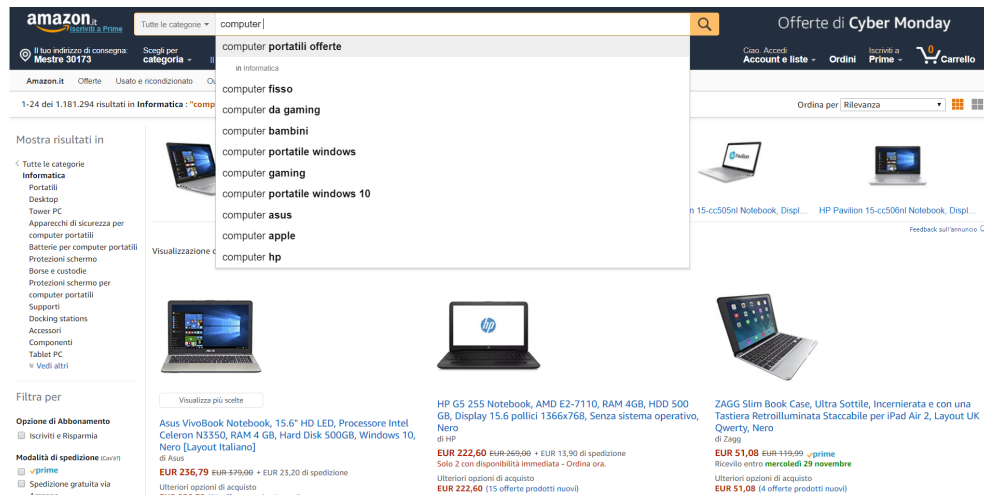


Figura 2.1: Esempi di funzionalità di ricerca offerte dal sito di Amazon

I motori di ricerca hanno assunto un ruolo fondamentale per gli utenti che navigano i siti web. Il genere di contatto che arriva da un motore di ricerca è particolarmente importante, in quanto quest'ultimo non rappresenta un utente passivo, bensì è un'entità attiva, pronta ad interagire con il suo utilizzatore. Durante una ricerca infatti, solitamente l'utente immette delle parole chiave, sperando di trovare ciò che sta cercando, con grandi aspettative derivanti dalla consultazione dei risultati.

Per aiutare l'utente a trovare ciò che sta cercando, i siti web devono però fare la loro parte: è necessario infatti che vengano curati con particolare attenzione alla visibilità elementi come testi e parole chiave.

Navigando i siti web, l'utente può trovare vari strumenti, messi a sua disposizione, che hanno lo scopo di aiutarlo a trovare quello che sta cercando. Tra i più importanti, troviamo:

- **Menù:** rappresentano delle categorizzazioni delle pagine; è necessario che queste siano ben categorizzate e le voci dei menù devono essere chiaramente distinguibili l'una dall'altra. Inoltre, è fortemente sconsigliato un numero elevato di voci e di sottomenù: l'utente deve infatti poter ritrovare l'informazione che cerca in un numero molto limitato di click. Per siti che contengono una grande mole di informazioni, contenute in un gran numero di pagine, questo strumento risulta dunque essere utile ma non sufficiente, in quanto non rappresenta per l'utente uno strumento attraverso il quale trovare agevolmente i contenuti di interesse in un sito contenente una grande mole di documenti;

- **Search Box:** l'utente deve inserire in un campo di testo una o più parole chiave. Eventualmente, è possibile disporre di una ricerca avanzata, specificando uno o più filtri. Qualsiasi etichetta o filtro che permetta di inserire criteri come “cerca solo la parola esatta” oppure “escludi dalla ricerca le seguenti parole” funziona solamente se l'utente è capace di esprimere in termini logici i criteri che ha in mente. Per essere efficace, è dunque di fondamentale importanza disporre di una ricerca che sia il più possibile in grado di comprendere il linguaggio naturale dell'utente;
- **Faceted Search:** questo strumento estende l'idea dei semplici filtri di ricerca, consentendo di ricercare elementi attraverso filtri multipli, comprendendo quindi più attributi per volta nella ricerca.

Lo scopo dello stage consisteva nell'analisi delle caratteristiche dei motori di ricerca [open source](#) nell'ambito dei siti web di tipo informativo e, in particolare, lo studio e il confronto delle funzionalità di possibile interesse per i siti web istituzionali delle Camere di Commercio. L'obiettivo da raggiungere, consiste nel cercare di migliorare l'esperienza di navigazione, da parte degli utenti, all'interno dei siti camerali, rendendo la navigazione il più possibile efficiente, efficace e semplice.

Per raggiungere questo obiettivo, in primo luogo mi è stato richiesto un approfondimento delle caratteristiche istituzionali dei siti web delle Camere di Commercio.

Successivamente, l'azienda richiedeva lo studio della tecnologia [Drupal](#) e dell'analisi delle potenzialità e specificità dei motori di ricerca [Sorl](#) e [ElasticSearch](#).

Veniva inoltre richiesto un prototipo di un sito web in tecnologia [Drupal](#), con i due motori di ricerca precedentemente citati.

Infine, era richiesta una relazione finale sulle potenzialità emerse nell'utilizzo dei due motori di ricerca.

2.2.2 Obiettivi posti dall'azienda

Le milestone

Il Piano di Lavoro, redatto assieme al tutor aziendale che mi ha seguito durante lo stage in InfoCamere S.C.p.A. , prevede, tra l'altro, una serie di [milestone](#); ad ognuna di esse, sono associati i prodotti sviluppati entro ogni corrispondente scadenza. Di seguito ne riporto l'elenco completo:

- Fine prima settimana: Ambiente di sviluppo configurato e funzionante (n.b. per ambiente di sviluppo si intende installazione in locale di un sito web [Drupal](#));
- Fine seconda settimana: Relazione con approfondimenti riguardanti [Sorl](#);
- Fine quarta settimana: Realizzazione prototipo di sito web in [Drupal](#) integrato con [Sorl](#);
- Fine quinta settimana: Relazione con approfondimenti riguardanti [ElasticSearch](#);
- Fine settima settimana: Realizzazione prototipo di sito web in [Drupal](#) integrato con [ElasticSearch](#);
- Fine ottava settimana: Relazione conclusiva e presentazione dell'elaborato al team di lavoro.



Figura 2.2: Milestone e metodologia di lavoro

Prodotti attesi

L'attività di stage prevedeva la produzione di un insieme di oggetti, frutto di tale attività. Di seguito, ne è riportato l'elenco:

- Documento: relazione sul motore di ricerca [Sori](#) che riporti le caratteristiche funzionali ed architetture della soluzione, pregi e difetti e possibile utilizzo all'interno del contesto InfoCamere;
- Documento: relazione sul motore di ricerca [ElasticSearch](#) che riporti le caratteristiche funzionali ed architetture della soluzione, pregi e difetti e possibile utilizzo all'interno del contesto InfoCamere;
- Documento: relazione finale di comparazione dei due motori di ricerca;
- Software: prototipo di sito web in tecnologia [Drupal](#) che utilizza motore di ricerca [Sori](#);
- Software: prototipo di sito web in tecnologia [Drupal](#) che utilizza motore di ricerca [ElasticSearch](#);
- Documento: relazione conclusiva (slide) sull'esperienza dell'uso dei prototipi, pregi e difetti nell'uso dei motori di ricerca in [Drupal](#) e comparazioni.

Priorità degli obiettivi aziendali

Gli obiettivi aziendali riguardanti lo stage possono essere suddivisi in:

- obiettivi minimi: vincolanti in quanto richieste primarie del committente;
- obiettivi massimi: non vincolanti o strettamente necessari, ma dal riconoscibile valore aggiunto.

Di seguito, vengono riportati gli obiettivi minimi e massimi stabiliti:

ID	Descrizione
Obiettivi obbligatori (min)	
min01	Analisi dei punti di forza e debolezza dei prodotti Sori ed ElasticSearch
min02	Realizzazione del prototipo in Drupal con le funzioni minime di ricerca
Obiettivi desiderabili e opzionali (max)	
max01	Comparazione dei due motori di ricerca esaminati con altri di riferimento nel mercato

ID	Descrizione
max02	Indicazioni su possibili interventi sui siti web istituzionali per quanto riguarda la user experience di navigazione, a seguito delle potenzialità espresse dai motori di ricerca

Tabella 2.1: Obiettivi dello stage

2.2.3 Vincoli

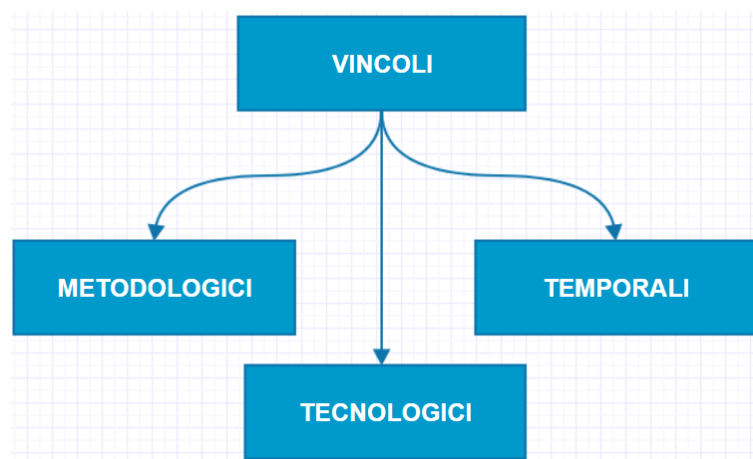


Figura 2.3: Tipologie di vincoli del progetto

Vincoli metodologici

In accordo con InfoCamere S.C.p.A. , ho svolto lo stage presso la sede di Padova (PD). In questo modo, ho avuto la possibilità di confrontarmi con programmatori più esperti ed essere supportato al meglio in caso di problematiche di sviluppo e gestione del progetto; lo svolgimento dello stage in azienda aveva inoltre lo scopo di favorire l'inserimento dello stagista nell'area di sviluppo aziendale. Ho avuto la possibilità, a cadenza quotidiana, di discutere con il tutor aziendale per qualsiasi tipo di problema legato allo stage. Settimanalmente, abbiamo fatto il punto della situazione sullo stato del lavoro, rivedendo, quando necessario, obiettivi settimanali o eventuali miglioramenti al progetto.

Nell'ultima settimana lavorativa, l'azienda ha inoltre richiesto una presentazione, esposta verbalmente e corredata da diapositive illustrative, sul lavoro operato durante lo stage e sulle relative conclusioni.

Altro vincolo legato alla metodologia, riguarda invece la tipologia di installazione dell'ambiente di sviluppo. In accordo con il tutor aziendale, ho deciso di installare l'intero ambiente di sviluppo localmente, così da poter intervenire agevolmente sull'intero sistema, senza la necessità di coinvolgere i sistemisti (o comunque altri reparti aziendali), per apportare modifiche a configurazioni delle tecnologie utilizzate, in modo tale da velocizzare i tempi di intervento.

Vincoli tecnologici

Per la gestione di versione dei prodotti dello stage, InfoCamere mi ha predisposto un repository [Git](#), nel quale ho versionato le varie istanze dei siti [Drupal](#) creati, assieme ai relativi [dump](#) dei database, oltre ai documenti scritti.

Per quanto concerne la tecnologia di comunicazione, mi sono invece state fornite delle credenziali di accesso, per poter accedere all'intranet aziendale e al client di posta elettronica utilizzato da InfoCamere S.C.p.A. ([Zimbra](#)), oltre a consentirmi l'accesso all'account del PC della postazione che mi è stata assegnata.

Per la presentazione finale fatta in azienda ho invece utilizzato un template aziendale dedicato alle presentazioni, realizzato con Microsoft PowerPoint.

Infine, le istanze [Drupal](#) create su [Acquia Dev Desktop 2](#) sono state realizzate in locale, seppur questa tecnologia fornisca la possibilità di ospitare le istanze sulla piattaforma [Acquia Cloud](#). Così facendo, è possibile non dipendere in alcun modo da tempistiche legate alla tipologia di un eventuale account su [Acquia Cloud](#). In futuro, sarà dunque possibile per l'azienda disporre liberamente delle istanze da me create.

Vincoli temporali

L'orario di lavoro era lo stesso previsto per il personale InfoCamere S.C.p.A. : dal Lunedì al Venerdì, con orari flessibili di un'ora dalle 08.00-09.00 alle 17.00-18.00 con un'ora di pausa pranzo.

L'azienda mi ha fornito un badge per accedere ai locali aziendali e mi è stato inoltre richiesto di timbrare quotidianamente il badge per registrare gli orari di entrata e di uscita. Inoltre, a cadenza quotidiana, il tutor aziendale registrava le attività svolte, la data e apponeva infine la propria firma.

Il numero complessivo di ore previsto dallo stage era pari a circa 300 ore, distribuite nell'arco di due mesi, con settimane lavorative di 40 ore ciascuna.

2.3 Obiettivi personali

Personalmente, ritengo che svolgere uno stage sia molto importante nell'odierno mercato del lavoro; questa opportunità porta infatti benefici, oltre che all'azienda, allo stagista stesso, sia dal punto di vista personale, sia da quello professionale.

Uno studente universitario apprende, nel corso del proprio percorso formativo scolastico, un gran numero di nozioni teoriche. Allo stesso tempo però, gli studenti hanno carenze sull'aspetto pratico del lavoro. Per poter sopperire a queste mancanze, permettendo allo studente di essere più competitivo nel mercato del lavoro, è di fondamentale importanza arricchire le conoscenze teoriche con esperienze professionali.

Lo stage permette allo studente di fare esperienza senza la necessità, da parte dell'azienda, di assumere personale privo di esperienza.

Oltre alle sopraccitate motivazioni, riporterò di seguito i motivi che mi hanno portato a scegliere, tra le tante aziende con le quali sono venuto a contatto grazie all'evento StageIT, InfoCamere S.C.p.A. .

- Motivazioni professionali: Essendo stata questa la mia prima esperienza lavorativa in un'azienda che lavora in ambito informatico, ho scelto InfoCamere S.C.p.A. , azienda con un gran numero di dipendenti e particolarmente strutturata, con la speranza di potermi confrontare con gente molto più esperta di me, con conoscenze più specifiche e capacità di problem solving differenti rispetto a

quelle che si possono trovare nel mondo universitario. Inoltre, ho ritenuto l'argomento dello stage di particolare interesse, in grado di fornirmi abilità molto richieste nell'attuale mondo del lavoro. Oltre a tutto questo, ritengo inoltre di fondamentale importanza ampliare la mia rete di contatti con professionisti che lavorino nel mio stesso campo d'interesse, utile per eventuali consigli professionali, raccomandazioni di lavoro, o più semplicemente per avere confronti costruttivi. Infine, l'esperienza porterà certamente ad un valore aggiunto al mio curriculum.

- **Motivazioni economiche:** A differenza di altre aziende con cui ho avuto un contatto, InfoCamere S.C.p.A. offriva un rimborso spese e buoni pasto. Ritengo che questo tipo di riconoscimento da parte dell'azienda sia un modo per incentivare lo stagista a svolgere al meglio il proprio lavoro, oltre ad agevolare lo studente, aiutandolo a sostenere le spese che deve sostenere.
- **Motivazioni personali:** Ritengo che lo stage sia il momento adatto per capire se il lavoro che ho scelto e la strada che ho intrapreso sia realmente ciò che voglio fare nella mia vita.

Capitolo 3

Resoconto dello stage

3.1 Individuazione dei motori di ricerca

In questa sezione motiverò la selezione di Solr e Elasticsearch come motori di ricerca da studiare.

3.2 Pianificazione

In questa sezione presenterò come è avvenuta la pianificazione dello stage.

3.3 I siti istituzionali delle Camere di Commercio

3.3.1 Funzionalità di ricerca attuali

Questa sezione presenterà la struttura dei siti camerali attualmente in produzione, ponendo l'accento sugli strumenti messi a disposizione all'utente per ritrovare i contenuti in esso presenti.

3.3.2 Possibile evoluzione

Questa sezione presenterà una possibile evoluzione dei siti camerali, contenente funzionalità di ricerca attualmente non presenti.

3.4 Ricerca nativa Drupal

3.4.1 Introduzione a Drupal

Qui verrà introdotto Drupal, spiegando cos'è e come funziona.

3.4.2 Ricerca di base e avanzata

Qui verranno presentate le principali funzionalità offerte dalla prima tipologia di ricerca nativa Drupal.

3.4.3 Ricerca con Search API

Qui verranno presentate le principali funzionalità offerte dalla seconda tipologia di ricerca nativa Drupal.

3.4.4 Considerazioni di Drupal nativo

Questa sezione conterrà conclusioni riguardanti le funzionalità di ricerca offerte globalmente dalla ricerca nativa Drupal.

3.5 Ricerca con Solr

3.5.1 Introduzione a Solr

Qui verrà introdotto Solr, spiegando cos'è e come funziona.

3.5.2 Principali funzionalità di ricerca

Qui verranno presentate le principali funzionalità di ricerca offerte dal motore di ricerca Solr, di possibile interesse per i siti camerali.

3.5.3 Integrazione con Drupal

Qui verranno discusse le funzionalità di ricerca derivanti dall'integrazione tra Solr e Drupal.

Apache Solr

Qui verranno discusse le funzionalità di ricerca derivanti dall'integrazione tra Solr e Drupal mediante il modulo Apache Solr Search.

Search API Solr

Qui verranno discusse le funzionalità di ricerca derivanti dall'integrazione tra Solr e Drupal mediante il modulo Search API Solr Search.

3.6 Ricerca con Elasticsearch

3.6.1 Introduzione a Elasticsearch

Qui verrà introdotto Elasticsearch, spiegando cos'è e come funziona.

3.6.2 Principali funzionalità di ricerca

Qui verranno presentate le principali funzionalità di ricerca offerte dal motore di ricerca Elasticsearch, di possibile interesse per i siti camerali.

3.6.3 Integrazione con Drupal

Qui verranno discusse le funzionalità di ricerca derivanti dall'integrazione tra Elasticsearch e Drupal.

Search API ElasticSearch

Qui verranno discusse le funzionalità di ricerca derivanti dall'integrazione tra ElasticSearch e Drupal mediante il modulo Search API ElasticSearch.

3.7 Considerazioni finali sui motori di ricerca esaminati

Questa sezione conterrà un confronto tra le principali funzionalità, possibilmente di interesse per l'azienda, offerte dalle tecnologie esaminate e quale di queste potrebbe essere la più adatta ai siti camerali.

Capitolo 4

Valutazione retrospettiva

4.1 Bilancio degli obiettivi

4.1.1 Aziendali

Questa sezione descriverà gli obiettivi aziendali soddisfatti, derivanti dallo stage.

4.1.2 Personali

Questa sezione descriverà gli obiettivi personali soddisfatti, derivanti dallo stage.

4.2 Conoscenze acquisite

Questa sezione descriverà le conoscenze acquisite derivanti dallo stage.

4.3 Mondo del lavoro e università a confronto

Questa sezione analizzerà il gap tra gli insegnamenti universitari e il mondo dello stage, specificatamente allo stage svolto.

Glossario

Acquia Cloud Servizio di hosting su piattaforma Acquia. 16, 25

Acquia Dev Desktop 2 E' un software per la realizzazione e la gestione di siti dinamici, che possono accrescere e mutare il proprio contenuto continuamente. Un **Content Management System** consente al committente del sito di occuparsi direttamente della sua gestione senza intermediari esterni. 16, 25

Apache Piattaforma server Web modulare largamente diffusa, in grado di operare su una grande varietà di sistemi operativi, tra cui UNIX/Linux, Microsoft Windows. 8, 25

API Insieme di procedure utilizzabili per interfacciarsi con un programma o un sistema informatico in modo standard. Spesso si intendono le librerie software disponibili in un certo linguaggio di programmazione. . 26, 27

CMS E' un software per la realizzazione e la gestione di siti dinamici, che possono accrescere e mutare il proprio contenuto continuamente. Un **Content Management System** consente al committente del sito di occuparsi direttamente della sua gestione senza intermediari esterni. 7, 25, 27

DAMP Acronimo che indica una piattaforma software per lo sviluppo di applicazioni web che prende il nome dalle iniziali dei componenti software con cui è realizzata. Le tecnologie contenute sono: **Drush**, **Apache**, **MySQL**, **PHP**. 8, 27

Drupal Drupal è un **Content Management System**, rilasciato sotto licenza **open source**, che permette la creazione di siti Internet, blog e portali, gallerie di immagini, forum di discussione, piattaforme intranet e molto altro. Essa è altresì un'applicazione completamente web based e può quindi essere utilizzata attraverso un semplice browser.
E' interamente sviluppato in **PHP** e utilizza come base di dati **MySQL** in modo nativo. v, 7–9, 13, 14, 16, 25

Drush Shell a riga di comando e interfaccia di scripting per **Drupal**. 25

Dump Il dump è un elemento di un database contenente un riepilogo della struttura delle tabelle del database medesimo e/o i relativi dati. 16, 25

ElasticSearch Piattaforma di ricerca **open source**, con capacità full text. E' un server di ricerca basato su **Java Lucene** e supporta architetture distribuite. Tutte le funzionalità sono nativamente esposte tramite interfaccia **RESTful**; le informazioni sono invece gestite come documenti **JSON**. v, 13, 14, 25

- Git** Sistema di controllo di versione distribuito e open source. 9, 16, 26
- Git Extensions** Interfaccia grafica per **Git** che consente l'utilizzo di **Git** senza dover ricorrere alla riga di comando. 9, 26
- HTTP** Formato adatto all'interscambio di dati fra applicazioni client-server. 26, 27
- Java** Linguaggio di programmazione ad alto livello, orientato agli oggetti e a tipizzazione statica, specificatamente progettato per essere il più possibile indipendente dalla piattaforma di esecuzione. 26
- Java Lucene API** gratuita ed **open source** per il reperimento di informazioni, inizialmente implementata in **Java**. 25, 26
- JSON** Formato adatto all'interscambio di dati fra applicazioni client-server. 25–27
- Milestone** Data temporale che indica il raggiungimento di determinati obiettivi intermedi nello svolgimento di un progetto. 13, 26
- Open source** Software di cui i detentori dei diritti rendono pubblico il codice sorgente, permettendo ad altri programmatori di apportarvi modifiche. Questo meccanismo è regolato tramite l'applicazione di apposite licenze d'uso. v, 7, 9, 13, 25, 26
- MySQL** Database relazionale largamente diffuso, composto da un client a riga di comando e un server. 7, 8, 25, 26
- PHP** Linguaggio di scripting interpretato. 7, 8, 25, 26
- REST** Stile architetturale che offre la possibilità di manipolare rappresentazioni testuali di risorse Web utilizzando un set predefinito di operazioni. 26
- RESTful** Le applicazioni basate su **REST**, si definiscono RESTful e utilizzano le richieste **HTTP** per inviare i dati (creazione e/o aggiornamento), effettuare query, modificare e cancellare i dati. In definitiva, **REST** utilizza **HTTP** per tutte e quattro le operazioni CRUD (Create / Read / Update / Delete). 25, 26
- Servlet** Oggetti scritti in linguaggio **Java** che operano all'interno di un server web oppure un server per applicazioni, permettendo la creazione di web applications. 26
- Solr** Piattaforma di ricerca **open source**. E' scritto in **Java** e viene eseguito come server di ricerca full text indipendente all'interno di un contenitore **Servlet**. Solr usa la libreria di ricerca **Java Lucene** per la ricerca e l'indicizzazione full text e mette a disposizione chiamate **REST** come ad esempio **HTTP**/**JSON** e **XML API** che rendono semplice la comunicazione. v, 13, 14, 26
- XML** Metalinguaggio che consente la rappresentazione di documenti e dati strutturati su supporto digitale. 26, 27
- Zimbra** Client di posta elettronica **open source**. 9, 16, 26

Acronimi

API Application Programming Interface. 25

CMS Content Management System. 25

DAMP Drush Apache MySQL PHP. 25

HTTP HyperText Transfer Protocol. 26

JSON JavaScript Object Notation. 26

XML eXtensible Markup Language. 26

Bibliografia

Siti web consultati

Displocazione delle Camere di Commercio Italiane e relative sedi distaccate. URL: <https://www.infocamere.it/noi-e-le-camere-di-commerico> (cit. a p. 3).

Esempio di attuale utilizzo del registro imprese. URL: <http://www.registroimprese.it/> (cit. a p. 2).

Nuovo istanza di un sito Drupal. URL: <https://www.drupal.org/documentation/install/beginners> (cit. a p. 8).

Servizi telematici offerti dal SUAP di Venezia. URL: <http://www.impresainungiorno.gov.it/web/guest/comune?codCatastale=L736> (cit. a p. 4).