

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

Daniel De Gaspari: *Analisi di motori di ricerca Open Source per siti web informativi*,
Tesi di laurea triennale, © Dicembre 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 [Sorl](#) 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	20
3.3 I siti istituzionali delle Camere di Commercio	21
3.3.1 Funzionalità di ricerca attuali	21
3.3.2 Possibile evoluzione	24
3.4 Ricerca nativa Drupal	28
3.4.1 Introduzione a Drupal	28
3.4.2 Ricerca di base e avanzata	29
3.4.3 Ricerca con Search API	30
3.4.4 Considerazioni di Drupal nativo	30
3.5 Ricerca con Solr	31
3.5.1 Introduzione a Solr	31

3.5.2	Principali funzionalità di ricerca	31
3.5.3	Integrazione con Drupal	34
3.6	Ricerca con ElasticSearch	34
3.6.1	Introduzione a ElasticSearch	34
3.6.2	Principali funzionalità di ricerca	35
3.6.3	Integrazione con Drupal	37
3.7	Considerazioni finali sui motori di ricerca esaminati	37
4	Valutazione retrospettiva	39
4.1	Bilancio degli obiettivi	39
4.1.1	Aziendali	39
4.1.2	PersonalI	39
4.2	Conoscenze acquisite	39
4.3	Mondo del lavoro e università a confronto	39
Glossario		41
Acronimi		45
Bibliografia		47

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>Displocazione 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. impresainun giorno.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
3.1	<i>Classifica dei motori di ricerca basata sulla popolarità.</i> URL: https://db-engines.com/en/ranking/search+engine	20
3.2	<i>Ricerca del termine "Conferenze" nell'attuale sito istituzionale camerale di Verona.</i> URL: https://www.vr.camcom.it/it/search/node	22
3.3	<i>Ricerca del termine "Conferenza" nell'attuale sito istituzionale camerale di Verona.</i> URL: https://www.vr.camcom.it/it/search/node	23
3.4	Funzionalità di completamento automatico di Google	24
3.5	Esempio funzionalità "Did you mean...?" in Google	25
3.6	Esempio funzionalità di evidenziamento, nei risultati, del termine ricercato	25
3.7	Esempio funzionalità di ricerca su file offerto da Google	26
3.8	Esempio funzionalità di filtri multipli (facets) presente su Amazon	26
3.9	Possibile evoluzione, mediante aggiunta di funzionalità di ricerca, del sito informativo Camerale di Verona	27
3.10	Utilizzo del motore da parte di un sito Camerale	28
3.11	Ricerca di base e avanzata, offerta dei moduli presenti nel core di Drupal	29
3.12	Logo ricerca Solr	31
3.13	Diagramma delle componenti principali di Solr	33

3.14 Logo ricerca ElasticSearch	34
3.15 Diagramma delle componenti principali di ElasticSearch	36
3.16 Principali funzionalità di possibile interesse per i siti informativi Camerali e relativo supporto da parte dei motori di ricerca esaminati	38

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.

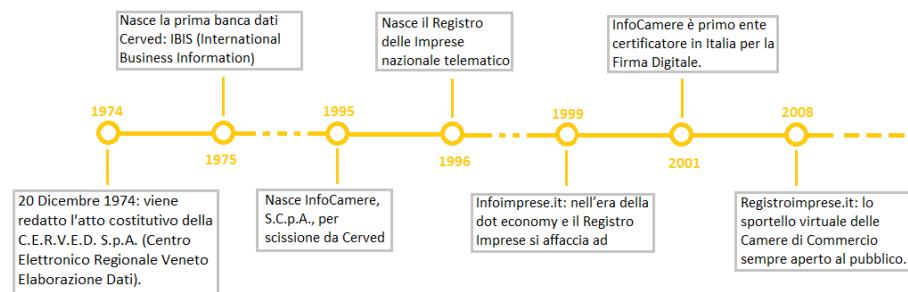


Figura 1.1: Storyline InfoCamere

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.

TIPOLOGIA	PREZZO
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 регистрационный сайт>

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, influendo 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 delle 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.

1.3. PROCESSI AZIENDALI

5

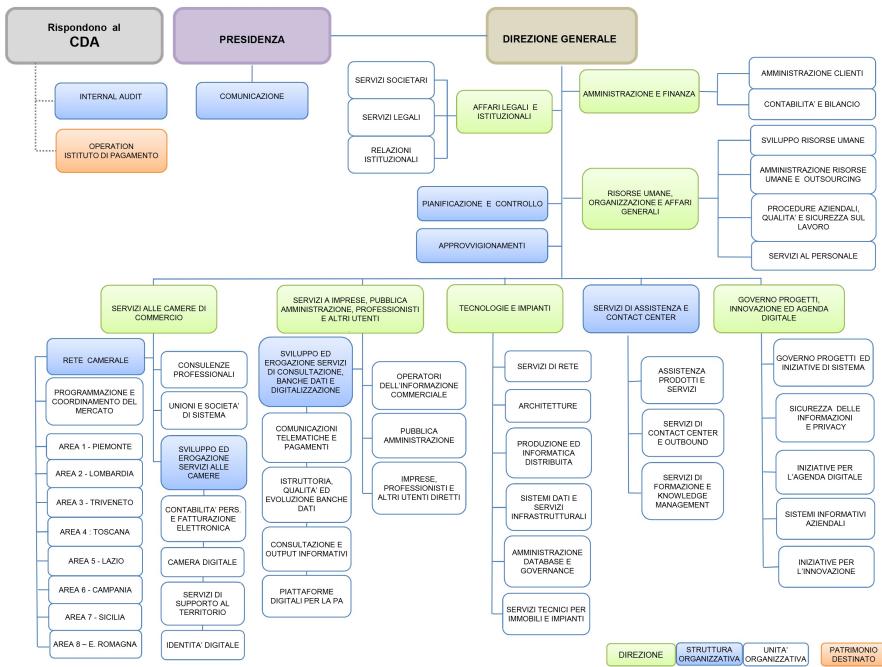


Figura 1.6: Organigramma aziendale

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.

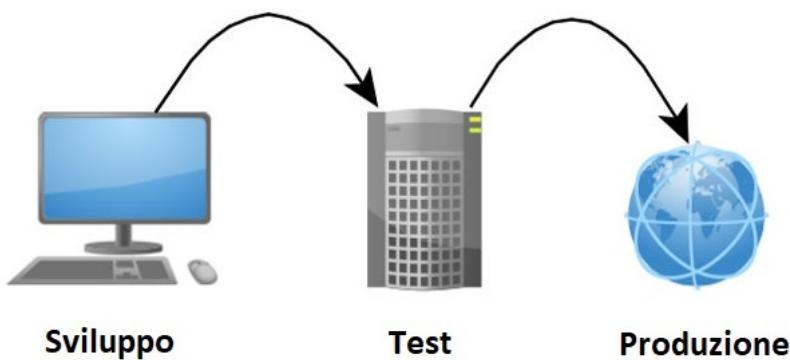


Figura 1.7: Ciclo di vita dei siti web

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. È interamente sviluppato in [PHP](#) e utilizza come base di dati [MySQL](#) in modo nativo.

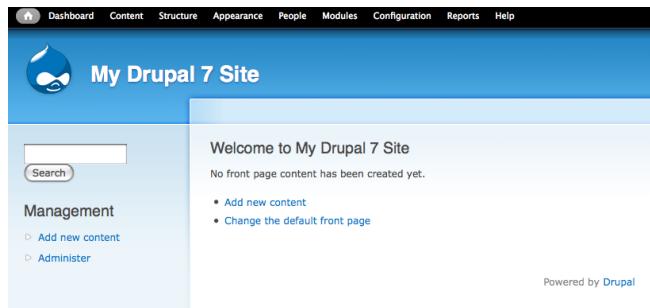


Figura 1.9: *Nuovo istanza di un sito Drupal.* URL: <https://www.drupal.org/documentation/install/beginners>

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 chiavi 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.

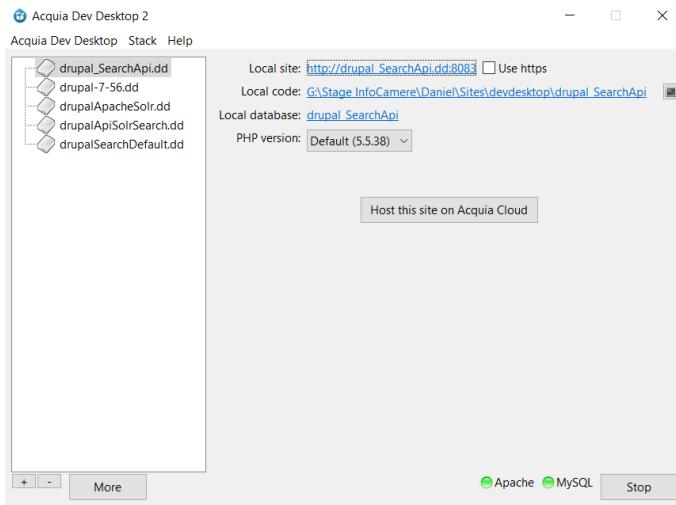


Figura 1.10: Pannello di controllo di Acquia Dev Desktop 2

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.

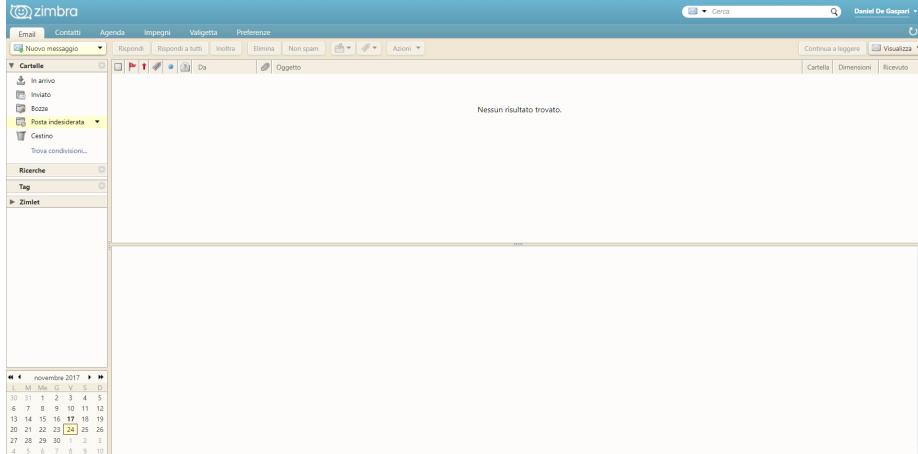


Figura 1.11: Client di posta elettronica Zimbra

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 aziendalemente. 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

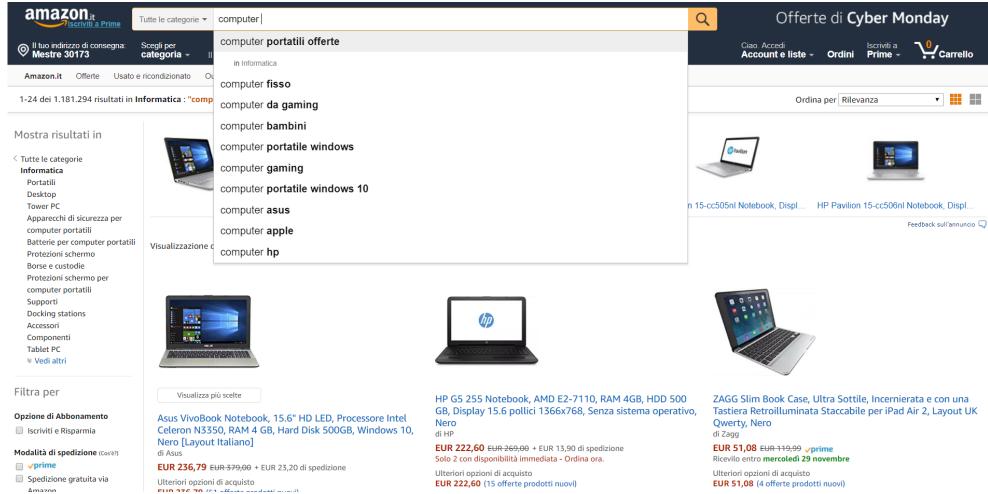


Figura 2.1: Esempi di funzionalità di ricerca offerte dal sito di Amazon

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 [Sorl](#) che riporti le caratteristiche funzionali ed architetturali 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 architetturali 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 [Sorl](#);
- 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 Sorl 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

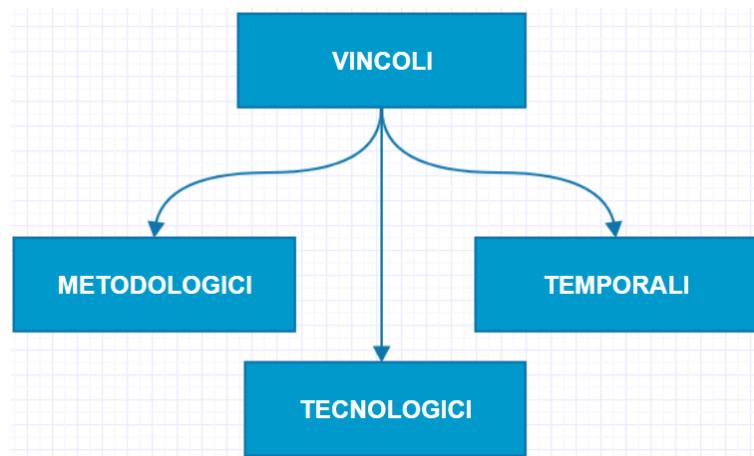


Figura 2.3: Tipologie di vincoli del progetto

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 illustrate, 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 soppiare 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. offre 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

Ad oggi, sono disponibili in rete un gran numero di motori di ricerca, sia proprietari, sia [open source](#). Questi, mettono a disposizione funzionalità di ricerca più o meno avanzate e, a seconda di vari fattori come possono essere, ad esempio, la tipologia e la quantità di dati che vogliamo indicizzare, rendendoli ricercabili agli utenti, possiamo optare per l'uno piuttosto che per l'altro.

La scelta del motore di ricerca più adatto alle specifiche esigenze, non è questione affatto banale; molte volte esistono varie soluzioni in grado di modellare bene il problema che dobbiamo affrontare, simili tra loro, che differiscono però per alcuni aspetti chiave. Focalizzando l'attenzione sulle esigenze dettate dai siti informativi Camerali, possiamo già applicare un primo filtro sui motori di ricerca da utilizzare: motori di ricerca proprietari *vs* motori di ricerca [open source](#).

Un motore di ricerca proprietario rappresenta indubbiamente un onere per l'azienda e, oltre a ciò, quest'ultima non ha il pieno controllo sulla destinazione finale dei dati; queste questioni non si hanno invece in un sistema [open source](#), nel quale il codice sorgente è liberamente modificabile e analizzabile, avendo dunque il pieno controllo sull'intero sistema.

Nella figura 3.1 è presente una classifica dei motori di ricerca, aggiornata a Novembre 2017, calcolata su parametri che rappresentano la popolarità del sito (maggiori informazioni sui punteggi assegnati sono disponibili al seguente indirizzo: https://db-engines.com/en/ranking_definition). Da questa tabella, notiamo come in testa alla classifica siano presenti tre motori di ricerca che hanno un ampio margine di distacco, in termini di punteggio ad essi attribuito, rispetto alle tecnologie che li precedono. Di questi, solamente i primi due ([Sorl](#) e [ElasticSearch](#)) sono alternative [open source](#).

Oltre a ciò, per entrambi i motori di ricerca appena citati, sono disponibili [moduli Drupal](#) che rendono possibile la semplice integrazione tra i motori di ricerca e il [Content Management System](#) di interesse.

Le motivazioni appena descritte, hanno così portato alla selezione di [Sorl](#) e [ElasticSearch](#) come motori di ricerca da studiare nell'ambito dello stage.

16 systems in ranking, November 2017										
Rank			DBMS	Database Model	Score			Nov 2017	Oct 2017	Nov 2016
Nov 2017	Oct 2017	Nov 2016			Nov 2017	Oct 2017	Nov 2016			
1.	1.	1.	Elasticsearch 	Search engine	119.41	-0.82	+16.84			
2.	2.	2.	Solr	Search engine	69.16	-1.97	+0.80			
3.	3.	3.	Splunk	Search engine	64.87	+0.51	+10.14			
4.	4.	4.	MarkLogic	Multi-model 	11.55	-0.26	+1.33			
5.	5.	5.	Sphinx	Search engine	5.88	-0.14	-1.11			
6.	6.	↑ 8.	Microsoft Azure Search	Search engine	3.88	+0.20	+1.96			
7.	↑ 8.	↑ 9.	Algolia	Search engine	2.78	+0.14	+1.23			
8.	↓ 7.	↓ 6.	Google Search Appliance	Search engine	2.76	-0.06	+0.14			
9.	9.	↓ 7.	Amazon CloudSearch	Search engine	2.36	-0.04	+0.10			
10.	10.	↑ 12.	CrateDB	Multi-model 	0.66	-0.08	+0.43			
11.	11.	↓ 10.	Xapian	Search engine	0.57	-0.02	+0.06			
12.	12.	↑ 14.	SearchBlox	Search engine	0.24	-0.01	+0.13			
13.	13.	↑ 16.	Exorbyte	Search engine	0.06	-0.02	+0.06			
14.	↑ 15.	↑ 16.	searchxml	Multi-model 	0.04	+0.04	+0.04			
15.	↓ 14.	15.	DBSight	Search engine	0.03	-0.01	+0.02			
16.	↓ 15.	↓ 11.	Indica	Search engine	0.00	±0.00	-0.36			

Figura 3.1: *Classifica dei motori di ricerca basata sulla popolarità.* URL: <https://db-engines.com/en/ranking/search+engine>

3.2 Pianificazione

Durante la stesura del Piano di Lavoro, assieme al tutor aziendale, ho concordato le attività principali da svolgere durante il periodo di stage, della durata di circa 300 ore, previsto dal mio corso di laurea. Nel documento sopracitato, sono inoltre contenute le *milestone* pianificate, così come descritto nella sezione 2.2.2, sotto alla voce "Le *milestone*".

Oltre al resoconto quotidiano con il tutor aziendale, al completamento di ogni attività, abbiamo fatto il punto della situazione sullo stato di avanzamento del lavoro, rivedendo e integrando, quando necessario, gli obiettivi settimanali previsti dal Piano di Lavoro. Questa modalità di operare mi ha permesso di rispettare le attività pianificate fin dall'inizio, oltre ad ampliare lo studio delle funzionalità di ricerca offerte da *Drupal*, non inizialmente previste, necessarie però a fornire un quadro più dettagliato sull'intero ambito di lavoro. Nell'ultima settimana lavorativa, l'azienda ha inoltre richiesto una presentazione, esposta verbalmente e corredata da diapositive illustrate, sul lavoro operato durante lo stage e sulle relative conclusioni. Più nel dettaglio, nella presentazione ho presentato: scopo e motivazioni del lavoro svolto; come ho organizzato il mio lavoro; funzionalità offerte dalle tecnologie da me esaminate (di interesse per l'azienda); conclusione soggettiva su quale tecnologia sia più adatta nell'ambito dei siti camerali, con suggerimenti per migliorare le funzionalità di ricerca dei suddetti siti; benefici immediati e a lungo termine che il lavoro svolto durante lo stage ha apportato all'azienda; ulteriori ambiti da esplorare, relativamente a queste tecnologie; valutazione soggettiva, mia e del tutor aziendale, riguardo allo stage come esperienza.

3.3 I siti istituzionali delle Camere di Commercio

3.3.1 Funzionalità di ricerca attuali

Quando un utente visita un sito web, è fondamentale che le informazioni in esso contenute siano facilmente trovabili dall'utente. A tal fine, è necessario che un sito web informativo fornisca all'utente strumenti che lo aiutino a trovare agevolmente, in pochi click, quello che sta cercando. La ricerca interna, spesso presente nei siti web, è uno di questi strumenti ed è strettamente legato ai concetti di:

- **efficienza:** il numero di azioni da che l'utente deve compiere per trovare quello che sta cercando dev'essere molto limitato; inoltre, i tempi di risposta devono essere molto brevi, quasi impercettibili agli occhi dell'utente, stando dunque nell'ordine dei msec;
- **efficacia:** l'utente deve essere in grado di trovare sempre quello che cerca; lo strumento di ricerca deve cercare di avvicinarsi a comprendere quanto più possibile il linguaggio naturale con il quale l'utente comunica al sito web.

Il fine ultimo dello stage consiste nel cercare di migliorare le funzionalità di ricerca offerte agli utenti, degli attuali siti web istituzionali delle Camere di Commercio.

Prendiamo come esempio l'attuale sito istituzionale della Camera di Commercio di Verona (*Sito web istituzionale della Camera di Commercio di Verona. URL: <https://www.vr.camcom.it/it/search/node>*): come presentato nella [figura 3.2](#), a fronte della ricerca del termine “Conferenze”, lo strumento di ricerca interno non trova alcun risultato, a differenza della ricerca di “Conferenza”, presentato nella [figura 3.3](#), che produce invece alcuni risultati. Rispetto alla prima ricerca effettuata, abbiamo semplicemente cambiato la forma del termine ricercato, dal plurale della prima ricerca al singolare della seconda.

Da questo comportamento, possiamo dunque trarre la conclusione che la ricerca viene attuata sui termini esatti, ed è ben lontana dal comprendere i linguaggi naturali.

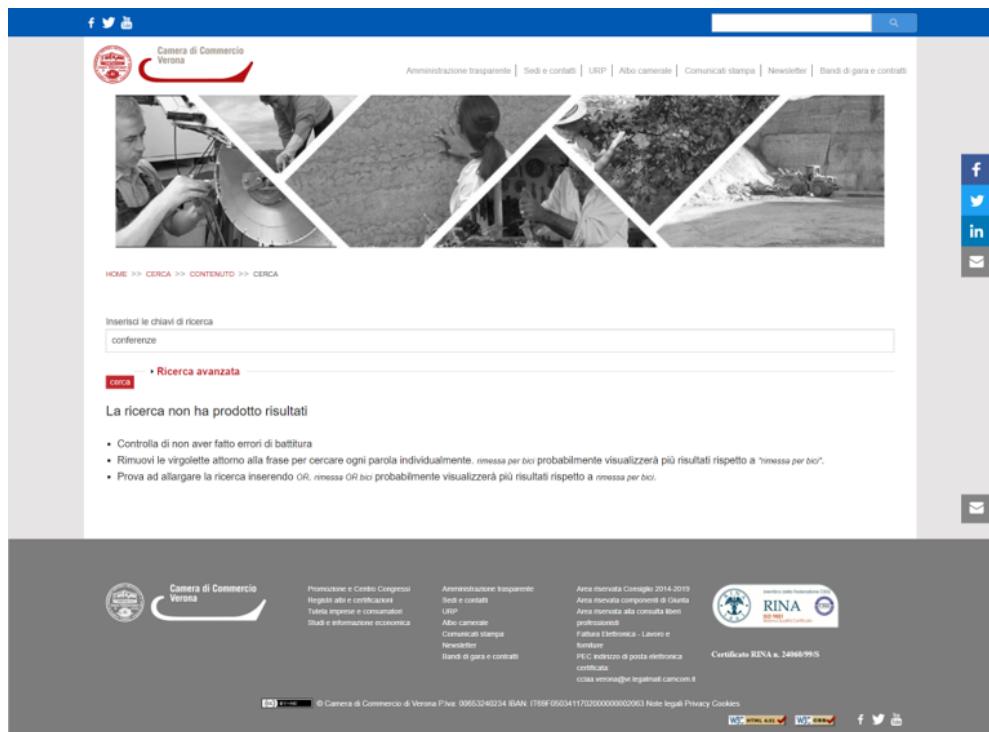


Figura 3.2: *Ricerca del termine "Conferenze" nell'attuale sito istituzionale camerale di Verona.* URL: <https://www.vr.camcom.it/it/search/node>

3.3. I SITI ISTITUZIONALI DELLE CAMERE DI COMMERCIO

23

Camera di Commercio Verona

Ammistrazione trasparente | Sedi e contatti | URP | Albo camerale | Comunicati stampa | Newsletter | Bandi di gara e contratti

Ricerca avanzata

Risultati della ricerca

NEWSLETTER DEL 12 SETTEMBRE 2017 N. 15
... un questionario via mail entro il 20 settembre. **Conferenza:** kiwi 2017/2018, le stime di produzione e commercializzazione ...
roberta doardo - 12/09/2017 - 17:48

NEWSLETTER DEL 20 LUGLIO 2016 - N. 13
... agosto 2016, esclusivamente per via telematica. **Conferenza** "Mele e pere 2016: previsioni produttive e situazioni di mercato" ... Consiliare del Comune di Zevio, è in programma l'annuale **conferenza** di presentazione delle previsioni produttive e situazioni di ...
roberta doardo - 20/07/2016 - 10:21

NEWSLETTER DEL 3 AGOSTO 2016 - N. 14
... provenienti da Paesi strategici per le imprese. **Conferenza** "Mele e pere 2016: previsioni produttive e situazioni di mercato" ... Consiliare del Comune di Zevio, è in programma l'annuale **conferenza** di presentazione delle previsioni produttive e situazioni di ...
paola.biondani - 03/08/2016 - 16:16

NEWSLETTER DEL 4 LUGLIO 2016 - N. 12
... di sfruttamento del marchio d'impresa all'estero. **Conferenza** "Mele e pere 2016: previsioni produttive e situazioni di mercato" ... Consiliare del Comune di Zevio, è in programma l'annuale **conferenza** di presentazione delle previsioni produttive e situazioni di ...
roberta doardo - 05/07/2016 - 15:02

26 settembre: conferenza Kiwi 2017/2018 stime di produzione e commercializzazione
... della Camera di Comercio di Verona per l'annuale **conferenza** "Kiwi 2017/2018: stime di produzione e commercializzazione". ... Tags kiwi Allegato Locandina **conferenza** kiwi 2017/2018 Martedì, Settembre 26, 2017 ...
lusella acabi - 26/09/2017 - 18:34

Conferenza: " Mele e pere 2017: previsioni produttive e situazioni di mercato"
... presso la Sala Consiliare del Comune di Zevio, l'annuale **conferenza** "Mele e pere 2017: previsioni produttive e situazioni di ... in particolare delle mele e delle pere. Durante la **conferenza** si discuterà anche dei problemi del comparto a livello nazionale, ...
paola.biondani - 14/08/2017 - 12:01

NEWSLETTER DEL 29 SETTEMBRE 2017 N. 16
... qualità ottima: questo il quadro che è emerso nella **conferenza** che si è tenuta in Camera di Comercio di Verona il 26 settembre. ...
roberta doardo - 29/09/2017 - 15:26

6 dicembre: conferenza "La cimice asiatica: una nuova minaccia per le produzioni agricole. Biologia, ecologia ed esperienze di difesa"
Data notizia Martedì, Novembre 7, 2017 Si terrà mercoledì 6 dicembre 2017, alle ore 9.30, presso l'Auditorium Domus Mercatorum della Camera di Comercio di Verona il convegno " La cimice asiatica: una nuova minaccia per le ...
roberta doardo - 20/11/2017 - 12:19

NEWSLETTER DEL 9 AGOSTO 2017 N. 13
... presso la Sala Consiliare del Comune di Zevio, l'annuale **conferenza** "Mele e pere 2017: previsioni produttive e situazioni di mercato" ...
paola.biondani - 09/08/2017 - 11:25

NEWSLETTER DEL 12 LUGLIO 2017 N. 12
... presso la Sala Consiliare del Comune di Zevio, l'annuale **conferenza** "Mele e pere 2017: previsioni produttive e situazioni di mercato" ...
paola.biondani - 12/07/2017 - 11:14

PROMOZIONE, PID E CENTRO CONFERENZE
Consigli e certificazioni
Tutela imprese e consumatori
Studi e informazione economica
Esposizioni artigiane: mostra dell'artigianato veneto in Gran Guardia dal 6 al 9 dicembre

Amministrazione trasparente
Sedi e contatti
URP
Albo camerale
Comunicati stampa
Newsletter
Bandi di gara e contratti

Area riservata Consiglio 2014-2019
Area riservata componenti di Giunta
Area riservata alle consolle liberi professionisti
Fattura Elettronica - Lavori e forniture
PEC andamento di posta elettronica certificata
posta.verona@legalmail.camcom.it

RINA
membro della Federazione CNA
Certificato RINA n. 24669/99/S

© Camera di Commercio di Verona P.Iva: 00653240234 IBAN: IT69F0503411702000000002063 Note legali Privacy Cookies

W3C HTML 4.01 W3C CSS3 f t g

Figura 3.3: *Ricerca del termine "Conferenza" nell'attuale sito istituzionale camerale di Verona.* URL: <https://www.vr.camcom.it/it/search/node>

3.3.2 Possibile evoluzione

Come ho illustrato nella sezione precedente, le funzionalità di ricerca attuali sono molto limitate, ben lontane dal fornire all'utente un'esperienza di navigazione incentrato sulla ricerca, con strumenti limitati nel numero e nella qualità delle funzionalità offerte. Come possiamo dunque migliorare la ricerca nei siti informativi Camerali? Per perseguire questo obiettivo ambizioso, pensiamo alle funzionalità di ricerca con cui gli utenti sono soliti interagire.

Google ad esempio, mette a disposizione diverse funzionalità che aiutano l'utente ad effettuare una ricerca che produca i risultati attesi, pur gestendo un'enorme quantità di dati. I siti di e-commerce utilizzano inoltre un altro strumento molto utile per una ricerca, che consente agli utenti di filtrare le ricerche, restringendo molto rapidamente i possibili contenuti di interesse: le facets, detti anche filtri multipli, che verranno presentati nel seguito di questo paragrafo. Alcune di queste funzionalità potrebbero essere prese come punto di partenza per migliorare la ricerca interna dei siti informativi Camerali. In particolare:

Completamento automatico

Questa caratteristica consente agli utenti di visualizzare una lista di termini suggeriti sulla base dei caratteri che hanno digitato

fino a quel momento; in questo modo, viene ridotto il numero di errori di digitazione della keywords da ricercare, consentendo così ricerche più vicine a quanto pensato dagli utenti, avvicinandosi in tal modo alle loro esigenze.

Ulteriore vantaggio è dato dalla riduzione del tempo richiesto agli utenti per effettuare una ricerca, in quanto se questa è presente tra le keywords suggerite, l'utente può risparmiare tempo semplicemente selezionando il suggerimento completo, piuttosto di dover digitare interamente i termini da ricercare, rendendo così la ricerca maggiormente efficiente in termini temporali.



Figura 3.4: Funzionalità di completamento automatico di Google

Paginazione e ordinamento

La funzionalità di paginazione permette di rendere le ricerche più rapide, restituendo un sottoinsieme dei risultati di ricerca. L'ordinamento consente invece di ordinare i risultati restituiti a seguito di una ricerca, sulla base di specifiche caratteristiche dei documenti indicizzati.

Controllo ortografico

Funzionalità indispensabile per comprendere quanto più possibile il linguaggio naturale con il quale l'utente comunica, il controllo ortografico può essere gestito come:

- Correzione automatica: viene eseguito automaticamente il controllo ortografico dei termini ricercati, sulla base del fatto che il termine contenente errori ortografici sia indicizzato;
- "Did you mean...?": vengono suggerite ricerche che potrebbero produrre risultati migliori; l'utente visualizzerà l'elenco di questi suggerimenti.

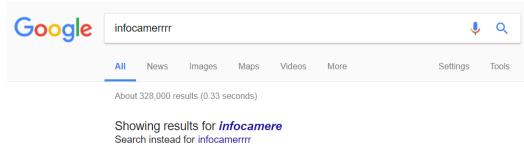


Figura 3.5: Esempio funzionalità "Did you mean..." in Google

Evidenziamento delle parole ricercate nei risultati

Quando vengono ricercati documenti che contengono una quantità significativa di testo, è bene fornire all'utente la possibilità di visualizzare sezioni specifiche di ogni documento, mettendo in risalto le sezioni dei risultati di ricerca contenenti i termini ricercati dagli utenti.



Figura 3.6: Esempio funzionalità di evidenziamento, nei risultati, del termine ricercato

Ricerca su file

L'utente può trovare, tra i risultati della ricerca che ha effettuato, documenti in vari formati (PDF, doc, ecc...) che contengono le keywords ricercate. Questa funzionalità risulta essere di particolare interesse nell'ambito dei siti informativi Camerali, data la grande mole di dati contenuti in documenti memorizzati in vari formati (prevalentemente PDF).

Facets

Prevalentemente utilizzati nei siti di e-commerce, l'utilizzo di questa funzionalità permette di fornire agli utenti strumenti utili a raffinare le proprie ricerche, ricorrendo

The screenshot shows a Google search results page with the following details:

- Search Query:** Come primo Prestatore di Servizi di Pagamento (PSP) abilitato al servizio @e bollo
- Number of Results:** Circa 4.620 risultati (0,66 secondi)
- Results:**
 - [PDF] COMUNICATO STAMPA - Infocamere
 - [PDF] Elenco dei Prestatori di Servizi di Pagamento abilitati all'incasso degli...
 - Infocamere, via ai pagamenti online tramite PagoPA - CorCom

Figura 3.7: Esempio funzionalità di ricerca su file offerto da Google

a una serie di filtri che consentono di ridurre rapidamente il numero dei risultati di ricerca, selezionando le caratteristiche di interesse di quanto ricercato. In questo modo, l'utente è in grado di costruire il proprio percorso di ricerca, consentendogli di effettuare ricerche più flessibili e che lo aiutino a trovare ciò che sta cercando con poche azioni. E' di fondamentale importanza che i contenuti siano ben categorizzati, così da permettere all'utente di trovare quanto cercato.

The screenshot shows an Amazon search results page for 'Notebook' with the following features:

- Left Sidebar Filters:**
 - Filtra per
 - Opzione di Abbonamento: Iscriviti e Risparmia
 - Modalità di spedizione (Corri!):
 - prime
 - Spedizione gratuita via Amazon
 - Giorno di consegna:
 - Ricevi oggi
 - Ricevi domani
 - Nuovi arrivi:
 - Ultimi 30 giorni
 - Ultimi 90 giorni
 - Disponibilità:
 - Includi non disponibili
 - Notebook RAM:
 - Fino a 2 GB
 - 3 GB
 - 4 GB
 - 6 GB
 - 8 GB
 - 12 GB
 - 16 GB e più
 - Velocità CPU:
 - Fino a 1.5 GHz
 - 1.6 - 2 GHz
 - 2.1 - 2.5 GHz
 - 2.5 GHz e più
- Right Side Product Listings:**
 - Asus VivoBook Notebook, 15.6" HD LED, Processore Intel Celeron N3350, RAM 4 GB, Hard Disk 500GB, Windows 10, Nero [Layout Italiano]**
di Asus
EUR 236,79 EUR 379,00 + EUR 23,20 di spedizione
Ulteriori opzioni di acquisto
EUR 236,79 (51 offerte prodotti nuovi)
★ ★ ★ ★ ★ 15
 - ZAGG Slim Book Case, Ultra Sottile, Incernierata e con una Tastiera Retroilluminata Staccabile per iPad Air 2, Layout UK Qwerty, Nero**
di Zagg
EUR 51,08 EUR 119,99 prime
Ricevilo entro mercoledì 29 novembre
Ulteriori opzioni di acquisto
EUR 51,08 (4 offerte prodotti nuovi)
★ ★ ★ ★ ★ 3

Figura 3.8: Esempio funzionalità di filtri multipli (facets) presente su Amazon

Importare le funzionalità nei siti informativi Camerali

Con l'obiettivo di migliorare la ricerca all'interno dei siti web informativi delle Camere di Commercio, in accordo con il tutor aziendale, abbiamo scelto di concentrare lo

studio delle funzionalità offerte dai motori di ricerca oggetto di analisi, principalmente sugli strumenti appena presentati.

Quello a cui vogliamo ambire, è presentato in un prototipo da me creato durante lo stage, contenente alcune delle funzionalità sopra esposti, visibile in Figura 3.9.

Il prototipo è stato creato con i medesimi dati contenuti nell'attuale sito web informativo della Camera di Comercio di Verona. Il termine ricercato presentato nella figura, è lo stesso della **ricerca effettuata nell'attuale sito istituzionale camerale di Verona**, ovvero "Conferenze", che non aveva prodotto alcun risultato. Notiamo invece che nel prototipo realizzato, la ricerca del medesimo termine sui medesimi contenuti ha invece prodotto alcuni risultati. Questo è stato possibile migliorando la funzionalità di ricerca, rendendola più flessibile e aumentando la capacità di comprensione del linguaggio naturale.

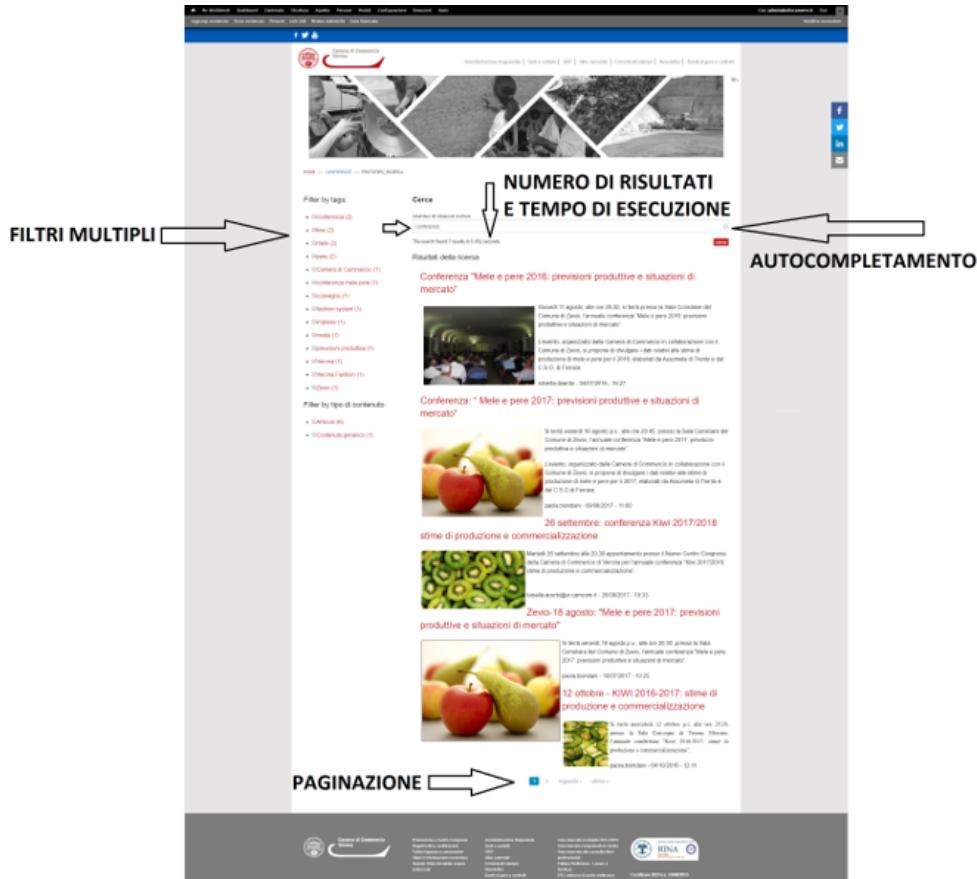


Figura 3.9: Possibile evoluzione, mediante aggiunta di funzionalità di ricerca, del sito informativo Camerale di Verona

3.4 Ricerca nativa Drupal

3.4.1 Introduzione a Drupal

Per arrivare a studiare le funzionalità offerte dai motori di ricerca [Solr](#) ed [ElasticSearch](#), mi sono confrontato con [Drupal](#), tecnologia attualmente utilizzata da InfoCamere S.C.p.A. per la creazione dei siti web informativi Camerali. Come viene presentato nella Figura 3.10, tale tecnologia è l’ambiente con il quale gli utenti comunicano. Le funzionalità di ricerca offerte dai siti creati in [Drupal](#) possono servirsi di moduli di ricerca internamente presenti all’ambiente, oppure utilizzare connettori in grado di usufruire di motori di ricerca esterni, come possono essere [Solr](#) ed [ElasticSearch](#).

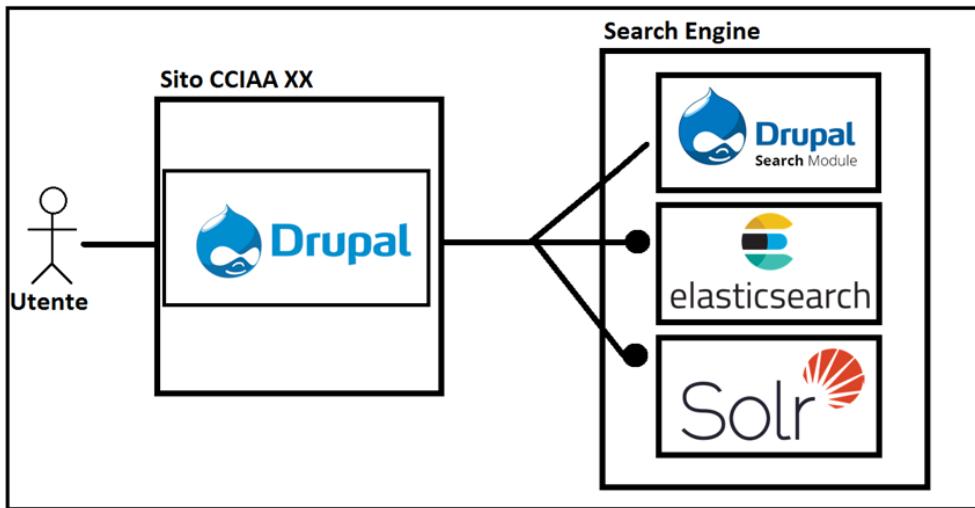


Figura 3.10: Utilizzo del motore da parte di un sito Camerale

Facciamo un passo indietro, e chiediamoci: cos’è Drupal? Questa tecnologia è un [Content Management System](#) che permette di creare siti web di vario tipo, rilasciato sotto licenza [open source](#) e interamente sviluppato in [PHP](#), supportando nativamente [MySQL](#) come base di dati [MySQL](#). Inoltre, rappresenta un’applicazione completamente web based, potendo dunque essere utilizzata attraverso un semplice browser. L’ambiente [Drupal](#) mette a disposizione [moduli](#) che consentono di aggiungere funzionalità di ricerca al sito che stiamo realizzando. In particolare, è disponibile una ricerca di base e avanzata e una ricerca che sfrutta il modulo Search API offerto da Drupal, che differiscono tra loro essenzialmente per le funzionalità offerte. Per ognuna di esse, ne ho analizzato le principali caratteristiche, focalizzando l’attenzione sulle funzionalità di possibile interesse per i siti web informativi camerali.

Scelta della versione Drupal

Sono disponibili diverse versioni [Drupal](#). Quella utilizzata per tutte le istanze create è la versione 7.56. La scelta deriva dalla volontà di partire da una istanza basilare, contenente solamente i moduli appartenenti al core di questa tecnologia. Esiste inoltre una versione più avanzata, ovvero Drupal 8.x; il motivo per cui la scelta è ricaduta sulla versione 7.x è dato dalla mancanza di moduli aggiornati alla versione 8. Per

questo motivo, viene dunque utilizzata una versione più vecchia, della quale sono però presenti un maggior numero di [moduli](#).

3.4.2 Ricerca di base e avanzata



Figura 3.11: Ricerca di base e avanzata, offerta dei moduli presenti nel core di Drupal

Modalità di operare

Per questa tipologia di ricerca, come per tutte quelle che seguono in questo capitolo, ho proceduto innanzitutto alla creazione di una nuova istanza [Drupal](#), dedicata alla tipologia di ricerca in esame. Dopo una iniziale configurazione dell'ambiente, ricercando e installando i moduli che mi fornissero le funzionalità di maggior interesse, ho proceduto allo studio delle funzionalità offerte.

Di seguito, riporto le conclusioni emerse dallo studio della ricerca di base e avanzata, resi disponibili dai moduli nativamente contenuti nel core di [Drupal](#).

Considerazioni sulla ricerca di base e avanzata

Una ricerca esaustiva prodotta con questa tecnologia richiede che l'utente conosca esattamente quello che sta ricercando; i contenuti hanno inoltre la necessità di essere fortemente atomizzati e specifici, mediante l'utilizzo di termini significativi, in modo tale da poter essere trovati dall'utente.

Nei siti informativi Camerali, questo limite può essere mitigato ricorrendo all'utilizzo di [tag](#), attraverso la creazione di [tassonomie](#), in modo da consentire all'utente una navigazione dei contenuti per tag. Così facendo, l'utente dovrà comunque conoscere con esattezza i termini chiave da ricercare e i contenuti dovranno essere etichettati in maniera corretta e specifica per i contenuti. Tutto ciò, porterebbe ad un rapido aumento del numero di tag, rendendo il sistema poco efficiente e fruibile.

Un'ulteriore difficoltà che incontra questo approccio è da ricercarsi nel carico di lavoro che il server sul quale il sito viene ospitato è in grado di gestire. Non essendoci infatti la possibilità di definire engine di ricerca esterni all'ambiente [Drupal](#), avendo quindi un database comune, utilizzato anche per effettuare le ricerche, un elevato quantitativo di interrogazioni al database potrebbe causare ritardi significativi o addirittura interruzioni dei servizi offerti dal sito.

Le ricerche di base e avanzata risultano dunque essere primitive, non adatte all'attuale stato di avanzamento tecnologico che si può osservare in vari siti informativi e non adatte a gestire siti di grandi dimensioni, dato lo scarso livello di performance e funzionalità offerte.

3.4.3 Ricerca con Search API

Modalità di operare

Per prima cosa, ho proceduto alla creazione di un'istanza [Drupal](#) dedicata alla ricerca di con [Search API](#). Oltre all'installazione di moduli di utilità, non significativi ai fini della ricerca ma che semplificano l'utilizzo dell'intero ambiente di sviluppo, ho installato il modulo [Search API](#).

Tale modulo fornisce un [Framework](#) per creare agevolmente ricerche nell'ambiente [Drupal](#), utilizzando qualunque tipologia di motore di ricerca. Prima di arrivare a studiare l'integrazione di questo modulo con i motori [Solr](#) e [ElasticSearch](#), ho sfruttato una ricerca che utilizzasse un [server](#) interno all'ambiente [Drupal](#), ritrovando ciò nel [modulo Search API Database](#). Di seguito, riporto le conclusioni emerse dallo studio della ricerca con [Search API](#) che sfrutta il modulo [Search API Database](#).

Considerazioni sulla ricerca con Search API Database

La ricerca [Search API](#) che utilizza [Search API Database](#) come [server](#) di ricerca, permette di effettuare ricerche in modo molto più flessibile rispetto a quanto offerto dalle ricerche di base e avanzata. Questa tecnologia permette un certo grado di complessità nelle stringhe ricercate, offrendo funzionalità molto più adatte alle aspettative degli utenti che navigano il web ai nostri giorni.

[Search API](#) non supporta tutte le funzionalità offerte dalla ricerca di base e avanzata: in particolare non sono presenti le funzionalità di ricerca che utilizzano clausole condizionali (and, or, not, ...).

Le funzionalità aggiuntive, non presenti nei [moduli](#) contenuti nel core di [Drupal](#), consentono però di eseguire ricerche più intelligenti, che aiutano l'utente a trovare contenuti ricercando termini anche non completamente esatti o incompleti.

Le ricerche risultano essere molto più user friendly rispetto a quelle rese disponibili dalla ricerca di base e avanzata, evidenziando un importante passo avanti, date le funzionalità offerte, permettendo agli utenti un margine di flessibilità nei termini ricercati, che non devono dunque conoscere esattamente il termine o il concetto che stanno ricercando.

Il carico di lavoro richiesto dal server sul quale il sito viene ospitato rappresenta un punto critico di questa tecnologia di ricerca: così come per la ricerca di base e avanzata, la ricerca [Search API](#) che utilizza [Search API Database](#), continua a lavorare con un unico database comune, utilizzato anche per effettuare ricerche. Un elevato quantitativo di richieste e l'impossibilità di esternalizzare l'engine di ricerca su macchine che lavorino in modo concorrente potrebbe dunque compromettere i tempi di attesa per ottenere i risultati di ricerca, che diventano critici su siti di grandi dimensioni e con un gran numero di contenuti.

3.4.4 Considerazioni di Drupal nativo

la ricerca nativa di [Drupal](#) permette un gran numero di funzionalità che potrebbero essere di interesse per i siti Camerali. Ulteriori [moduli](#) amplificano inoltre le funzionalità possibilmente utili disponibili mediante l'utilizzo di questa tipologia di ricerca, permettendo funzionalità completamento automatico, paginazione delle ricerche e ordinamento dei risultati di ricerca secondo filtri prestabiliti, "Did you mean...?", evidenziamento dei termini ricercati nei risultati, utilizzo di [moduli](#) per implementare le facets (=filtri multipli), creazione di più [index](#) di ricerca e possibilità di effettuare

ricerche nei contenuti testuali di alcuni tipi di files (PDF, Word, ecc...).

Questa tecnologia consente dunque di effettuare ricerche con un certo grado di complessità, offrendo funzionalità molto più adatte alle aspettative degli utenti che navigano il web nei nostri giorni.

La ricerca sulle [stem words](#) rappresenta un possibile punto debole di questa tecnologia, non essendo attualmente disponibile un modulo nella versione 7.x per la lingua italiana. Il carico di lavoro richiesto dal server sul quale il sito viene ospitato rappresenta un punto critico di questa tecnologia. In entrambe le tipologie di ricerca presentate per la ricerca nativa [Drupal](#), è presente un unico database comune. Un elevato quantitativo di richieste, unito all'impossibilità di esternalizzare l'engine di ricerca su macchine che lavorino in modo concorrente potrebbe dunque compromettere i tempi di attesa per ottenere i risultati di ricerca, che diventano critici su siti di grandi dimensioni e con un gran numero di contenuti.

3.5 Ricerca con Solr

3.5.1 Introduzione a Solr



Figura 3.12: Logo ricerca Solr

Solr rappresenta una piattaforma di ricerca [open source](#), scritta in [Java](#), che viene eseguita come server di ricerca full text indipendente, all'interno di un contenitore [Servlet](#). Utilizza [Java Lucene](#) come libreria per la ricerca e l'indicizzazione full text e mette a disposizione chiamate [REST](#) come ad esempio, comunicando in vari formati, tra i quali [JSON](#) e [XML](#), rendendo semplice e versatile la comunicazione.

Nonostante sia scritto in [Java](#), non è necessaria una conoscenza di tale linguaggio per utilizzare questa tecnologia; è inoltre possibile estendere la libreria di ricerca: in questo caso la conoscenza di [Java](#) risulterebbe, ovviamente, fondamentale.

Questa tecnologia mette a disposizione un gran numero di funzionalità che aiutano l'utilizzatore ad eseguire ricerche semplici e intuitive ma potenti al tempo stesso. E' importante notare che questa tecnologia espone [API](#) che offrono servizi [HTTP](#) di tipo [REST](#), senza fornire componenti grafici per la ricerca.

3.5.2 Principali funzionalità di ricerca

Modalità di operare

Dopo uno studio iniziale riguardante l'architettura ad alto livello della tecnologia, ho proceduto alla creazione e alla configurazione di una nuova istanza [Solr](#); la versione che

ho utilizzato è la 5.5.4. Successivamente, ne ho studiato le funzionalità di ricerca più significative offerte dalla tecnologia in esame, prestando attenzione alla sua integrazione con l'ambiente [Drupal](#).

Architettura ad alto livello

Per lo studio dell'architettura ad alto livello di questa tecnologia, mi sono servito di letteratura e materiale trovato in rete. Con le fonti trovate, ho individuato le seguenti componenti che vanno a comporre l'architettura di massima di [Sorl](#), presentate di seguito.

- **Request Handler:** Le richieste che arrivano al sistema [Sorl](#) vengono processate da questa componente e possono essere di interrogazione o di aggiornamento dell'[index](#);
- **Search Components:** Rappresenta l'insieme delle funzionalità di ricerca (faceting, highlighting, more like this, ecc...) offerte da Apache Solr. Le singole componenti di ricerca (search component) si registrano come gestori di ricerca (search handlers). E' inoltre possibile registrare più componenti in un singolo search handler;
- **Query Parser:** Il [Parser](#) di Apache Solr analizza le query passate a [Sorl](#), verificandone la correttezza sintattica. Successivamente, traduce le interrogazioni in un formato comprensibile a [Java Lucene](#);
- **Response Writer:** Componente che si occupa della generazione dell'output a seguito delle interrogazioni eseguite. Vengono supportati vari formati, tra cui [XML](#), [JSON](#), ecc...;
- **Analyzer/tokenizer:** Apache Solr analizza il contenuto e lo divide in [tokens](#) che verranno successivamente inviati a [Java Lucene](#). Un Analyzer esamina il testo dei campi del contenuto e genera un flusso di token. Successivamente, questo flusso di token viene suddiviso in singoli token da un [Tokenizer](#);
- **Update Request Processor:** Ogni richiesta di aggiornamento inviata ad Apache Solr passa attraverso una serie di processi, responsabili delle modifiche (eliminazioni, aggiunte di campi, ecc...), che insieme vengono definiti update request processor.

Considerazioni sulla ricerca Solr

Per ogni tipologia di funzionalità di interesse per i siti informativi Camerali, ho proceduto alla configurazione delle componenti in modo tale da creare esempi funzionanti che supportassero la funzionalità in esame. In particolare, ho potuto constatare che tutte le funzionalità di possibile interesse per InfoCamere S.C.p.A. , presentate in [3.3.2](#), vengono offerte e supportate da [Sorl](#).

Lo step successivo, consisteva nel verificare se tali funzionalità continuavano ad essere supportate integrando questo motore di ricerca con l'ambiente [Drupal](#), in quanto quest'ultimo avrebbe potuto agire da collo di bottiglia.

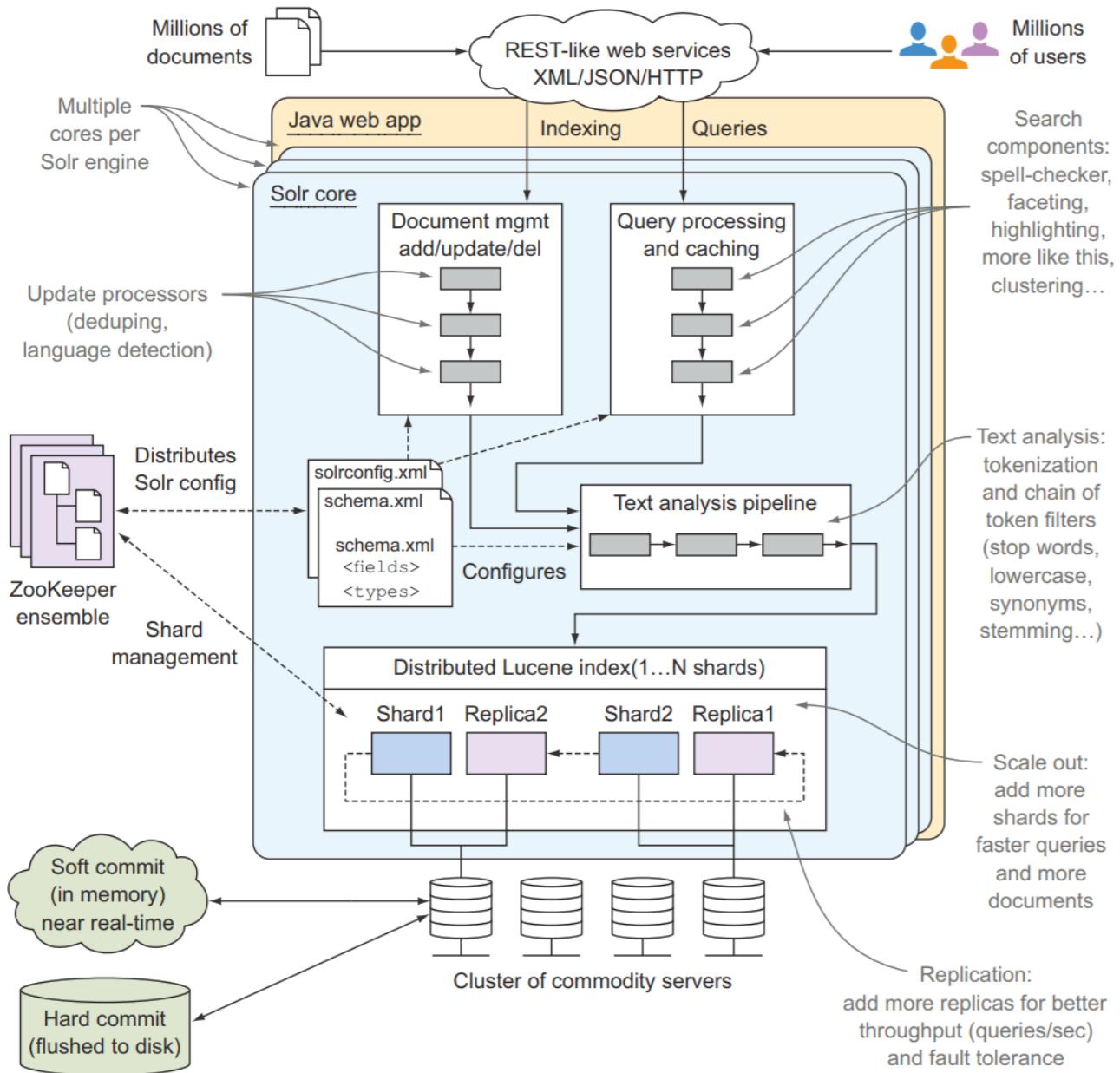


Figura 3.13: Diagramma delle componenti principali di Solr

3.5.3 Integrazione con Drupal

Le versioni di [Drupal](#) (7.56) e di [Sorl](#) (5.5.4) utilizzate permettono l'integrazione di queste due tecnologie in due differenti modalità, presentate di seguito.

Apache Solr

Apache Solr permette di integrare il motore di ricerca [Sorl](#) con l'ambiente [Drupal](#) mediante utilizzo di un apposito [modulo](#), denominato "Apache Solr". Questo modulo mette in comunicazione il modulo di ricerca "Search" offerto dal core di [Drupal](#) con il motore di ricerca.

Dopo aver creato e configurato un'apposita istanza [Drupal](#) che utilizzasse il suddetto modulo, ho potuto constatare che, pur utilizzando un potente motore di ricerca, non vengono supportate alcune delle funzionalità di possibile interesse per i siti Camerali, data l'assenza di [moduli](#) necessari all'integrazione della ricerca di base e avanzata con [Sorl](#). In particolare, un'istanza [Drupal](#) che utilizzi il modulo "Apache Solr" non permette l'utilizzo dei filtri multipli nè la creazione di più indici, utili per gestire tipologie di ricerca che differiscano per tipo di documento/entità ricercata, all'interno di un sito.

Search API Solr

Search API Solr permette di integrare il motore di ricerca [Sorl](#) con l'ambiente [Drupal](#) utilizzando il modulo "Search API Solr", che funge da connettore tra le due tecnologie, mettendo in comunicazione il modulo di ricerca [Search API](#) con il motore di ricerca. Dopo aver creato e configurato un'apposita istanza [Drupal](#) che utilizzasse il suddetto modulo, ho avuto modo di vedere che tutte le funzionalità di possibile interesse per i siti Camerali vengono ben supportate da questa tipologia di integrazione, risultando essere lo strumento di maggior interesse tra tutti quelli presentati fino ad ora.

3.6 Ricerca con ElasticSearch

3.6.1 Introduzione a ElasticSearch



Figura 3.14: Logo ricerca ElasticSearch

[ElasticSearch](#) è una piattaforma di ricerca [open source](#), con capacità di ricerca 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](#).

Questa tecnologia è divenuta molto popolare nell'ambito Big Data, negli ambienti enterprise e nel settore del cloud computing, date le funzionalità di ricerca, analisi e visualizzazione dei contenuti nei documenti [JSON](#), con interrogazioni che avvengono quasi in tempo reale.

Assieme a [Kibana](#) e [Logstash](#), produce una soluzione integrata, in grado di gestire grandi quantità di log in una forma centralizzata, facilmente ricercabile e che usa un'interfaccia grafica intuitiva; questi tre software vanno a produrre quello che viene definito "Elastic Stack" (ELK stack).

3.6.2 Principali funzionalità di ricerca

Modalità di operare

Dopo uno studio iniziale riguardante l'architettura ad alto livello della tecnologia, ho proceduto alla creazione e alla configurazione di una nuova istanza [ElasticSearch](#); la versione che ho utilizzato è la 5.6.4. Per semplificare l'utilizzo di questo strumento, ho utilizzato [cerebro](#), un tool che consente di avere una visione grafica degli indici di ricerca creati in [ElasticSearch](#) e che fornisce uno strumento per effettuare, in modo semplice, chiamate [REST](#), in modo tale da poter comunicare con il server di ricerca. Successivamente, ne ho studiato le funzionalità di ricerca più significative offerte dalla tecnologia in esame, prestando attenzione alla sua integrazione con l'ambiente [Drupal](#).

Architettura ad alto livello

L'architettura di [ElasticSearch](#) pone le sue fondamenta nel linguaggio [Java](#) e consente l'esecuzione di ricerche full text all'interno di un contenitore [Servlet](#). Così come accade per [Sorl](#), [ElasticSearch](#) utilizza [Java Lucene](#) come libreria per la ricerca. Qui le comunicazioni passano da un'interfaccia di tipo [RESTful](#) e le informazioni vengono gestite come documenti [JSON](#).

Una delle caratteristiche importanti di [ElasticSearch](#) riguardanti la gestione degli event log, è quella di essere prima di uno schema, rendendo in tal modo possibile, a seguito dell'inserimento di un qualsiasi documento in formato [JSON](#), l'indicizzazione di tutti i suoi campi, consentendo così di poter essere utilizzati per eseguire query di ricerca.

Nella letteratura e nel materiale ricercato in rete, non ho trovato una spiegazione chiara ed esaustiva sull'architettura ad alto livello di questa tecnologia. Assieme al tutor, partendo dall'architettura studiata per [Sorl](#), abbiamo dunque cercato di disegnare un'architettura di massima di questo motore di ricerca. Le componenti che abbiamo individuato sono le stesse di quelle presenti per [Sorl](#), presentate in precedenza in questo capitolo.

Considerazioni sulla ricerca ElasticSearch

Così come ho agito con lo studio di [Sorl](#), anche per [ElasticSearch](#), per ogni tipologia di funzionalità di interesse per i siti informativi Camerali, ho proceduto alla configurazione delle componenti in modo tale da creare esempi funzionanti che supportassero la funzionalità in esame. Anche in questo caso, ho potuto constatare che tutte le funzionalità di possibile interesse per InfoCamere S.C.p.A. , presentate in [3.3.2](#), sono ben supportate da [ElasticSearch](#).

Lo step successivo, consisteva nel verificare se tali funzionalità continuavano ad essere supportate integrando questo motore di ricerca con l'ambiente [Drupal](#), in quanto quest'ultimo avrebbe potuto agire da collo di bottiglia.

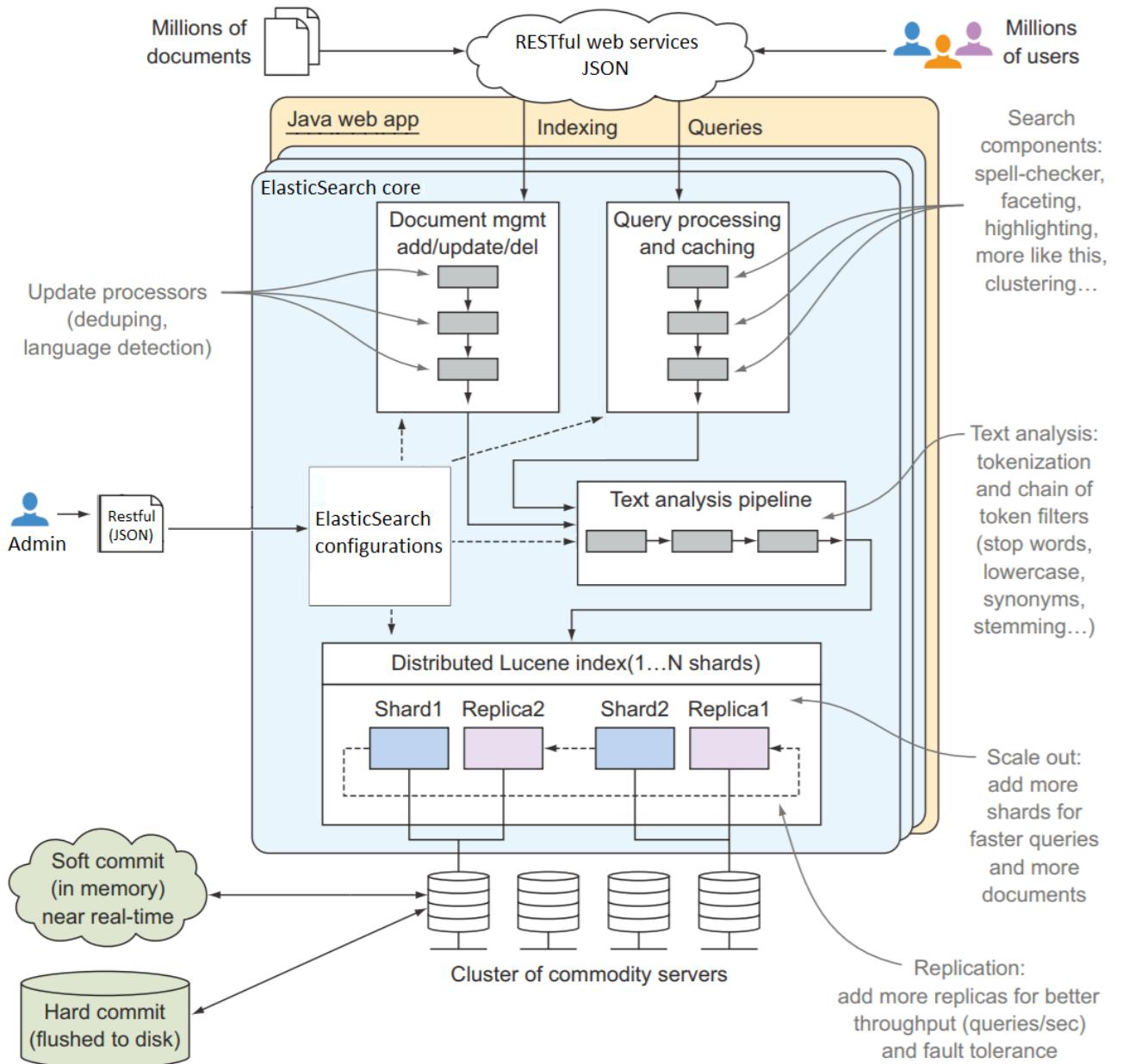


Figura 3.15: Diagramma delle componenti principali di Elasticsearch

3.6.3 Integrazione con Drupal

Le versioni di [Drupal](#) (7.56) e di [ElasticSearch](#) (5.6.4) utilizzate permettono l'integrazione di queste due tecnologie in una sola modalità.

Search API ElasticSearch

Search API Solr permette di integrare il motore di ricerca [Solr](#) con l'ambiente [Drupal](#) utilizzando il modulo "Search API ElasticSearch", che funge da connettore tra le due tecnologie, mettendo in comunicazione il modulo di ricerca [Search API](#) con il server di ricerca in esame.

Dopo aver creato e configurato un'apposita istanza [Drupal](#) che utilizzasse il suddetto modulo, ho avuto modo di vedere che tutte le funzionalità di possibile interesse per i siti Camerali, risultando essere dunque una valida alternativa a quanto offerto dalla ricerca con Search API Solr.

3.7 Considerazioni finali sui motori di ricerca esaminati

Dallo studio sulle funzionalità supportate dalle tecnologie di ricerca esaminate, ho prodotto una tabella riassuntiva, riportata di seguito, che illustri le funzionalità rese disponibili dalle varie tecnologie. E' bene sottolineare che tecnologie diverse possono supportare più o meno bene le singole funzionalità; un esempio per comprendere meglio questo concetto emerge dalla funzionalità di ricerca sui files: il numero di tipologie di file supportati è molto maggiore negli ambienti che utilizzano i motori di ricerca [Solr](#) e [ElasticSearch](#), a differenza invece della ricerca nativa, in grado di supportare un insieme molto limitato di estensioni, tra i quali è però presente il formato PDF, che risulta essere quello di maggior interesse per l'azienda.

Dal confronto emerge che le uniche due tecnologie in grado di supportare tutte le funzionalità individuate sono la ricerca con Search API Solr e la ricerca con Search API ElasticSearch.

Tra tutte queste, ho cercato di individuare quale delle due fosse più adatta per i siti web informativi Camerali realizzati da InfoCamere S.C.p.A.. Considerando la soluzione attuale presente in alcuni degli attuali siti in produzione, ovvero la ricerca nativa "core", la soluzione più adatta che ho individuato consiste nell'evolvere la ricerca attuale alla ricerca nativa con Search API, in grado di supportare tutte le funzionalità di possibile interesse, eccetto lo stemming (per la ricerca delle [stem words](#)). In un secondo momento, sarà possibile utilizzare un connettore per completare la lista delle funzionalità offerte, migliorando inoltre quelle presenti in Search API, mediante l'utilizzo di un motore di ricerca [Solr](#), che presenta il vantaggio di essere già conosciuto e utilizzato in azienda, dell'alternativa [ElasticSearch](#), o eventualmente di qualche altra nuova tecnologia che prenda piede nel mondo dei motori di ricerca.

	Ricerca Nativa		Ricerca Solr		Ricerca ElasticSearch
	Core	Search API	Apache Solr	Search API Solr	Search API ElasticSearch
Flessibilità nelle stringhe ricercate	✗	✓	✓	✓	✓
Stemming	✗	✗	✓	✓	✓
Autocomplete	✓	✓	✓	✓	✓
Paginazione e ordinamento	✓	✓	✓	✓	✓
Suggester (Did you mean...?)	✗	✓	✓	✓	✓
Highlighting	✓	✓	✓	✓	✓
Facets	✗	✓	✗	✓	✓
Multiple index	✗	✓	✗	✓	✓
Ricerca sui files	✓	✓	✓	✓	✓

Figura 3.16: Principali funzionalità di possibile interesse per i siti informativi Camerali e relativo supporto da parte dei motori di ricerca esaminati

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](#), [41](#)

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](#), [41](#)

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](#), [41](#)

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. . [31](#), [42](#), [43](#), [45](#)

Backend Insieme delle applicazioni e dei programmi con cui l'utente non interagisce direttamente ma che sono essenziali al funzionamento del sistema. [41](#), [43](#)

Cerebro Strumento per l'amministrazione web e monitoraggio che semplificano l'utilizzo di [ElasticSearch](#). [35](#)

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](#), [19](#), [28](#), [41](#), [45](#)

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](#), [45](#)

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](#), [19](#), [20](#), [28–32](#), [34](#), [35](#), [37](#), [41–43](#)

Drush Shell a riga di comando e interfaccia di scripting per [Drupal](#). [41](#)

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](#), [41](#)

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](#), [19](#), [28](#), [30](#), [34](#), [35](#), [37](#), [42](#), [43](#)

Framework Insieme di classi cooperanti che forniscono lo scheletro di un'applicazione riusabile per uno specifico dominio applicativo. Delinea l'architettura delle applicazioni in cui viene usato. [30](#), [42](#), [43](#)

Git Sistema di controllo di versione distribuito e [open source](#). [9](#), [16](#), [42](#)

Git Extensions Interfaccia grafica per [Git](#) che consente l'utilizzo di [Git](#) senza dover ricorrere alla riga di comando. [9](#), [42](#)

HTTP Formato adatto all'interscambio di dati fra applicazioni client-server. [31](#), [43](#), [45](#)

Index Oggetto di configurazione per l'indicizzazione dei dati che determina come e quali dati sono indicizzati a seconda delle configurazioni che assume. Tiene inoltre traccia di quali elementi devono ancora essere indicizzati (o re-indicizzati a seguito di aggiornamenti). Devono risiedere su un server per poter essere utilizzati. E' possibile ad esempio avere un "Node index" per indicizzare gli elementi di tipo nodo. L'indice è indipendente dalla meccanica interna utilizzata dai motori di ricerca. [30](#), [32](#), [42](#), [43](#)

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. [31](#), [35](#), [42](#), [43](#)

Java Lucene API gratuita ed [open source](#) per il reperimento di informazioni, inizialmente implementata in Java. [31](#), [32](#), [34](#), [35](#), [42](#), [43](#)

JSON Formato adatto all'interscambio di dati fra applicazioni client-server. [31](#), [32](#), [34](#), [35](#), [42](#), [43](#), [45](#)

Kibana Plugin [open source](#) che consente di visualizzare i dati per la tecnologia [ElasticSearch](#). Permette inoltre agli utenti di creare vari tipi di grafici basati su grandi volumi di dati. Assieme a [ElasticSearch](#) e [Logstash](#) compone quello che viene definito "Elastic Stack" (ELK stack). [35](#), [42](#)

Logstash Strumento per la gestione di eventi e log, fornendo un valido sistema per la raccolta, gestione e memorizzazione di attività legate alla ricerca. Assieme a [ElasticSearch](#) e [Kibana](#) compone quello che viene definito "Elastic Stack" (ELK stack). [35](#), [42](#)

Milestone Data temporale che indica il raggiungimento di determinati obiettivi intermedi nello svolgimento di un progetto. [13](#), [20](#), [42](#)

Modulo Collezione di funzioni che forniscono funzionalità aggiuntive alle istanze [Drupal](#). [19](#), [28–30](#), [34](#), [42](#), [43](#)

Open source Software di cui i detentori dei diritti rendono pubblico il codice sorgente, permettendo ad altri programmatore di apportarvi modifiche. Questo meccanismo è regolato tramite l'applicazione di apposite licenze d'uso. [v](#), [7](#), [9](#), [13](#), [19](#), [28](#), [31](#), [34](#), [41–44](#)

Parser Programma che analizza un file, verificandone la correttezza sintattica rispetto a una data grammatica. [32](#), [43](#)

MySQL Database relazionale largamente diffuso, composto da un client a riga di comando e un server. [7](#), [8](#), [28](#), [41](#), [43](#)

PHP Linguaggio di scripting interpretato. [7](#), [8](#), [28](#), [41](#), [43](#)

REST Stile architetturale che offre la possibilità di manipolare rappresentazioni testuali di risorse Web utilizzando un set predefinito di operazioni. [31](#), [35](#), [43](#)

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). [34](#), [35](#), [42](#), [43](#)

Search API Framework che consente di eseguire ricerche su qualunque tipo di entità conosciuta a **Drupal**, dando la possibilità di utilizzare qualunque tipo di motore di ricerca. E' un modulo di **Drupal**. [30](#), [34](#), [37](#), [43](#)

Search API Database modulo Drupal che fornisce un **backend**, semplice da utilizzare ed economico rispetto a sistemi più avanzati come **Solr** e **ElasticSearch**, per **Search API**; prevede un database per consentire l'indicizzazione dei dati. [30](#), [43](#)

Server Si occupa dell'effettiva indicizzazione dei dati. Può avere un numero arbitrario di **index** ad esso associati. [30](#), [43](#)

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. [31](#), [35](#), [43](#)

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](#), [19](#), [28](#), [30–32](#), [34](#), [35](#), [37](#), [43](#)

Stem Words Rappresentano la radice di un termine; da queste possono derivare le desinenze del termine radice, ovvero l'elemento finale variabile di una parola. [31](#), [37](#), [43](#)

Tag Parola chiave il cui intento è indicare senza troppi dettagli il contenuto della pagina a cui si riferiscono. [29](#), [43](#)

Tassonomia Nel suo significato più generale, rappresenta la disciplina della classificazione. Nell'ambito **Drupal**, rappresenta uno strumento (derivante da un apposito **modulo**) che consente di organizzare e catalogare i contenuti del sito. [29](#), [43](#)

Token Blocco di testo categorizzato. [32](#), [44](#)

Tokenizer Trasforma uno stream di testo in un elenco di **Token**. [32](#), [44](#)

XML Metalinguaggio che consente la rappresentazione di documenti e dati strutturati su supporto digitale. [31](#), [32](#), [43](#), [45](#)

Zimbra Client di posta elettronica open source. [9](#), [16](#), [44](#)

Acronimi

API Application Programming Interface. 41

CMS Content Management System. 41

DAMP Drush Apache MySQL PHP. 41

HTTP HyperText Transfer Protocol. 42

JSON JavaScript Object Notation. 42

XML eXtensible Markup Language. 44

Bibliografia

Siti web consultati

Classifica dei motori di ricerca basata sulla popolarità. URL: <https://db-engines.com/en/ranking/search+engine> (cit. a p. 20).

Dislocazione delle Camere di Commercio Italiane e relative sedi distaccate. URL: <https://www.infocamere.it/noi-e-le-camere-di-commercio> (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).

Ricerca del termine "Conferenza" nell'attuale sito istituzionale camerale di Verona. URL: <https://www.vr.camcom.it/it/search/node> (cit. a p. 23).

Ricerca del termine "Conferenze" nell'attuale sito istituzionale camerale di Verona. URL: <https://www.vr.camcom.it/it/search/node> (cit. a p. 22).

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

Sito web istituzionale della Camera di Commercio di Verona. URL: <https://www.vr.camcom.it/it/search/node> (cit. a p. 21).