GARE PUBBLICHE

Analisi funzionale

Sommario

[1. Obiettivo e ambito del documento 3](#_Toc485817552)

[2. Premessa 3](#_Toc485817553)

[3. Entità di business 8](#_Toc485817554)

[3.1. PUBBLICAZIONE 8](#_Toc485817555)

[3.2. BANDO 8](#_Toc485817556)

[4. Stati 9](#_Toc485817557)

[4.1. PUBBLICAZIONE 9](#_Toc485817558)

[4.2. BANDO 9](#_Toc485817559)

[5. Modello Entità-Relazione 10](#_Toc485817560)

[5.1. PUBBLICAZIONE 11](#_Toc485817561)

[5.2. BANDO 12](#_Toc485817562)

[5.3. EXPREG 13](#_Toc485817563)

[5.4. UTENTE 14](#_Toc485817564)

[6. Processi 15](#_Toc485817565)

[6.1. RICERCA NUOVA PUBBLICAZIONE 15](#_Toc485817566)

[6.2. SCARICA PUBBLICAZIONE 15](#_Toc485817567)

[6.3. PARSIFICA BANDI 16](#_Toc485817568)

[6.4. NOTIFICA BANDI 16](#_Toc485817569)

[7. Dettagli Realizzazione Tecnica 16](#_Toc485817570)

[7.1. ARCHITETTURA 16](#_Toc485817571)

[7.2. GESTIONE STATI 17](#_Toc485817572)

[7.3. STACK TECNOLOGICO 17](#_Toc485817573)

# Obiettivo e ambito del documento

###### Scopo del presente documento è definire le specifiche tecniche del progetto per il sistema “Gare Pubbliche”, un software web-based che consente la fruizione dei dati relativi alle gare pubbliche esposte sulla Gazzetta Ufficiale italiana ed europea.

# Premessa

###### L'obiettivo del progetto è di raccogliere e memorizzare i **bandi delle gare pubbliche** pubblicati dalla **Gazzetta Ufficiale**, per poi notificare gli utenti registrati al sistema della presenza degli stessi, appena sono disponibili; gli utenti possono esprimere delle preferenze sui bandi a cui sono interessati e sono notificati soltanto se i bandi rispecchiano le preferenze.

###### La Gazzetta è un documento ufficiale pubblicato dalla Republica Italiana con cadenza al massimo giornaliera e viene esposta in una pagina web al seguente indirizzo: <http://www.gazzettaufficiale.it/> (**Home Page della Gazzetta**). Le sue pubblicazioni sono identificate da data e un numero, che si azzera con la fine dell'anno. Una pubblicazione espone, oltre a decreti, sentenze e regolamenti, diverse decine di bandi di vario tipo, offerti da enti anche molto diversi quali Regioni, Province, Comuni, Ministeri...

###### Le **pubblicazioni** dei bandi pubblici della Gazzetta degli ultimi 30 giorni sono esposte nel sito web della Gazzetta Ufficiale, al seguente URL: <http://www.gazzettaufficiale.it/30giorni/contratti> (**Pagina delle Pubblicazioni di Bandi pubblici**).

### E' possibile raggiungere la pagina delle pubblicazioni dei bandi pubblici nella sezione "Gazzetta Ufficiale" esposta nella home page, in particolare scegliendo "5a Serie Speciale - Contratti Pubblici".

###### I **bandi** rappresentano delle gare pubbliche, cioè appalti di lavori o incarichi pubblici offerti dallo Stato e liberamente fruibili da privati o associazioni, previo concorso. Le informazioni principali che riguardano i bandi sono CIG (Codice Identificativo Gara) e oggetto della gara.

### Per ogni nuovo bando disponibile si memorizzano in modo persistente queste informazioni, per poi notificare gli utenti interessati.

###### Il sistema prevede un meccanismo di registrazione dei nuovi **utenti** con relativo username e password. Ogni utente imposta alcuni parametri che rappresentano le sue **preferenze** riguardo a certi bandi.

### Gli utenti possono definire gli argomenti che interessano tramite alcune espressioni da ricercare nel testo del bando. Si definiscono le espressioni che devono comparire nel testo e quelle che invece non devono essere presenti; se poi il bando contiene una o più espressioni della prima categoria ma nessuna della seconda, allora l'utente è notificato.

###### Quando dei nuovi bandi sono esposti sulla Gazzetta sono scaricati e memorizzati; il sistema sceglie quali sono gli utenti registrati interessati ai nuovi bandi: analizza il testo del bando e cerca se sono presenti o no le parole chiave definite dagli utenti, poi li notifica.

### Le **notifiche** avvengono tramite servizi Internet: un sito web e eventualmente posta elettronica, se l'utente lo ha specificato. In fase di registrazione l'utente definisce pertanto anche se desidera ricevere le notifiche per e-mail, oltre che visualizzarle sul sito web.

Pagina delle Pubblicazioni di Bandi pubblici

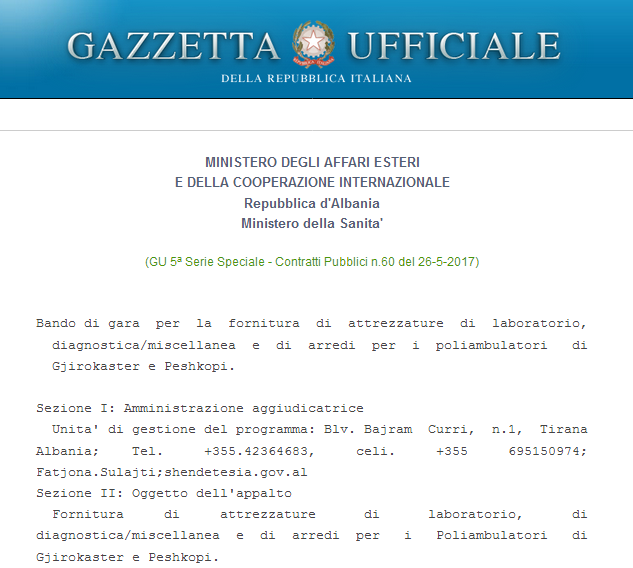
Home Page della Gazzetta Ufficiale

###### Gazzetta_Pub.pngNella pagina delle pubblicazioni dei bandi è possibile scegliere una pubblicazione e consultarne la relativa pagina che contiene l'elenco dei bandi e i link agli stessi. Da questa schermata si può inoltre accedere all'archivio completo di tutte le pubblicazioni.

Pagina di una particolare Pubblicazione

###### Gazzetta_Una_Pub.pngDalla pagina di una pubblicazione si possono consultare i diversi bandi esposti (**Pagina di una particolare Pubblicazione**). Le informazioni che riguardano il nome del richiedente, la data di scadenza e il codice identificativo di un bando sono già disponibili in questa pagina.

###### Scegliendo un particolare bando all'interno della pagina della pubblicazione, si va alla pagina del bando, dove è esposto tutto il testo del bando.

Pagina di un particolare bando

# Entità di business

###### Le Entità di Business coinvolte dal progetto sono **PUBBLICAZIONE** e **BANDO**.

## PUBBLICAZIONE

###### Le pubblicazioni della Gazzetta ufficiale sono di vari tipi e riguardano argomenti molto diversi: leggi, decreti, comunicati, sentenze, direttive europee e gare pubbliche. Analizzando la home page della Gazzetta si possono scegliere le diverse pubblicazioni:

###### Serie Generale : decreti legge e comunicati

### Prima serie speciale: sentenze ed ordinanze della Corte Costituzionale

### ...

### **Quinta serie speciale: contratti pubblici**

###### La quinta serie speciale riguarda i contratti pubblici: è la categoria di pubblicazioni che contiene i bandi di gare pubbliche, oggetto dell'applicazione.

###### Una pubblicazione è al suo interno divisa in sezioni, ognuna delle quali raccoglie i bandi a seconda dei vari sotto-argomenti che trattano. Nella quinta serie speciale (contratti pubblici) sono presenti quattro sezioni:

###### Avvisi e bandi di gara: la sezione che contiene i nuovi appalti pubblici disponibili per la gara.

### Avvisi esiti di gara: contiene i bandi di gara pubblicati in date precedenti a questa pubblicazione, e ora non più disponibili perchè assegnati a qualche ente che ha vinto la gara.

### Avvisi di annullamento, differimento e rettifica relativi a bandi e procedure di gara.

### Aste immobiliari ed altre procedure ad evidenza pubblica.

###### Le quattro sezioni non sono sempre tutte esposte in ogni uscita: se in una pubblicazione non sono presenti bandi che riguardano una certa sezione, questa è omessa.

###### La struttura della pagina delle pubblicazioni è stata già descritta: per raggiungere tale pagina è necessario andare sulla home page della Gazzetta, seguire il link a **pagina** **delle** **pubblicazioni,** scegliere una fra le varie serie presenti e infine una particolare pubblicazione. La pagina della pubblicazione espone diversi bandi corredati da: tipologia richiedente, nome del richiedente, scadenza, link al bando di gara e codice del bando. In alcuni casi è anche possibile trovare il CIG.

## BANDO

### All'interno di una pubblicazione della quinta serie speciale (contratti pubblici), la prima sezione è l'unica rilevante per il progetto in quanto contiene i nuovi bandi di gare pubbliche disponibili e non ancora aggiudicati.

###### Analizzando questi bandi si nota che purtroppo non rispettano una struttura uniforme e definita. Si possono però classificare in due o tre forme, a seconda di come sono stati redatti:

###### La prima forma è la meglio strutturata: c'è un titolo (che può contenere anche il CIG) e il documento è suddiviso in sezioni e sottosezioni quasi sempre uguali. In questo caso risulta facile trovare l'oggetto del bando, il suo codice e le altre informazioni utili.

###### Esempio:

###### Bando di gara n. 6756896 - CIG 7095200842

###### SEZIONE I: AMMINISTRAZIONE AGGIUDICATRICE: .....

###### SEZIONE II:

###### II.1) Descrizione.

###### II.1.1) Denominazione

###### II.1.4) Breve descrizione / oggetto dell'appalto

###### II.1.5) Breve descrizione dell'appalto o degli acquisti:

###### SEZIONE VI: altre informazioni

###### VI.3) Informazioni complementari: Codice identificativo della gara (CIG): ...

###### Una seconda forma strutturata, ma suddivisa soltanto in sezioni, in cui può o meno comparire il titolo della sezione. Anche in questa serie di documenti risulta comunque non troppo complesso ricercare l'oggetto dell'appalto e il CIG (se è presente!)

###### Esempio:

###### Bando di gara - Servizi - CIG 7095999B9C

###### SEZIONE II: OGGETTO DELL'APPALTO:... (CIG 7093944BC5)

###### SEZIONE II: Oggetto

###### SEZIONE II:

###### L'ultima forma raccoglie tutti i documenti non strutturati e quindi quelli che presentano una descrizione unica, non divisa in sezioni. Ovviamente la difficoltà nel ritrovare le informazioni è maggiore e spesso i documenti di questo tipo non presentano neanche il CIG.

###### Nell prime due forme le informazioni di principale interesse si possono reperire nel titolo e/o nella seconda sezione; per quanto riguarda l'ultima forma non strutturata, bisogna ricercare all'interno di tutto il bando senza avere particolari indicazioni su dove si trovano CIG e oggetto.

### Una volta analizzati questi documenti ne si estrae quindi il CIG e l'oggetto del bando, per poi memorizzarli in modo persistente.

# Stati

###### Gli stati riguardano le entità Pubblicazione e Bando e sono necessari per la gestione interna dei dati e dei processi che coinvolgono le due entità.

### Gli stati sono modificati nelle diverse fasi del processo ed esprimono un'informazione utile alle altre fasi, che si comportano in modo diverso a seconda del valore dello stato.

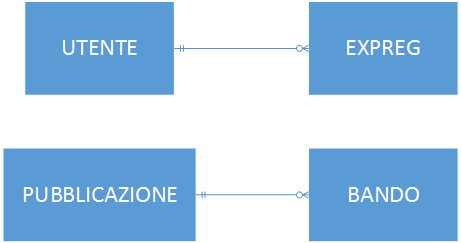
## PUBBLICAZIONE

|  |  |  |
| --- | --- | --- |
| Nome Stato | Valore | Commenti |
| STATO\_PUBBLICAZIONE | DA\_SCARICARE | Nuova pubblicazione rilevata: pronta da scaricare |
| SCARICATA | La pubblicazione è stata scaricata e i suoi bandi sono pronti ad essere parsificati |

## BANDO

|  |  |  |
| --- | --- | --- |
| Nome Stato | Valore | Commenti |
| STATO\_BANDO | DA\_PARSIFICARE | Il bando è stato scaricato tutto ed è pronto per essere parsificato |
| PARSIFICATO | Si è conclusa l'analisi del bando ed è pronto per essere pubblicato |

# Modello Entità-Relazione



## PUBBLICAZIONE

###### Permette di modellare le pubblicazioni della Gazzetta Ufficiale

|  |  |  |  |
| --- | --- | --- | --- |
| Nome Proprietà | Tipo | Null / Extra | Commenti |
| CD\_PUBBLICAZIONE | BIGINT | PRIMARY, AUTO\_INCREMENT | Identificativo della pubblicazione. Popolato automaticamente |
| DT\_INSERIMENTO | DATETIME |  | Data del'inserimento della pubblicazione nel database |
| NM\_PUBBLICAZIONE | INTEGER |  | Numero della pubblicazione |
| STATO | ENUM |  | DA\_SCARICARE, SCARICATA |
| URL | VARCHAR |  | Indirizzo URL della pubblicazione |

Tabella 1 - Entità PUBBLICAZIONE

## BANDO

###### L'entità bando rappresenta un bando di gara pubblica contenuto in una pubblicazione della Gazetta.

|  |  |  |  |
| --- | --- | --- | --- |
| Nome Proprietà | Tipo | Null / Extra | Commenti |
| CD\_BANDO | BIGINT | PRIMARY, AUTO\_INCREMENT | Identificativo del bando, generato in automatico |
| CD\_ESTERNO | VARCHAR (250) | UNIQUE | Codice del bando che si trova nella pubblicazione (TX17BFC9332) |
| CD\_PUBBLICAZIONE | BIGINT | REFERENCES PUBBLICAZIONE (CD\_PUBBLICAZIONE) | Rappresenta la relazione fra BANDO e PUBBLICAZIONE |
| CIG | VARCHAR (250) | Null | Codice Identificativo Gara, riportato sul bando |
| TIPO | VARCHAR (50) | Null | Bando di Gara / Estratto di bando di gara / Bando di gara a procedura aperta |
| TIPORICHIEDENTE | VARCHAR (100) | Null | Ministeri / Regioni / Province / ... |
| NM\_RICHIEDENTE | VARCHAR | Null | Il nome dell'ente richiedente (Associazione o privato) |
| SCADENZA | DATE | Null | La data di scadenza del bando, riportata sullo stesso |
| OGGETTO | VARCHAR | Null | Oggetto del bando |
| TESTO | TEXT | Null | Tutto il testo integrale del bando |
| URL | VARCHAR |  | Indirizzo web identificativo del bando |
| STATO | ENUM |  | DA\_PARSIFICARE, PARSIFICATO |
| DT\_INSERIMENTO | DATETIME |  | Data di inserimento del bando nel database, che corrisponde alla data di inserimento della pubblicazione |

Tabella 2 - Entità BANDO

###### Come identificativo del bando si può sfruttare il codice già assegnato dalla Gazzetta ai vari bandi (Esempio: TX17BFC9367) che si suppone essere univoco.

## EXPREG

###### L'entità gestisce le espressioni regolari associate a ciascun utente, che rappresentano i parametri con cui l'utente esprime il suo interesse in determinati tipi di bando.

|  |  |  |  |
| --- | --- | --- | --- |
| Nome Proprietà | Tipo | Null / Extra | Commenti |
| CD\_EXPREG | BIGINT | PRIMARY, AUTO\_INCREMENT | Identificativo dell'espressione, assegnato in automatico. |
| CD\_UTENTE | NUMBER | REFERENCES UTENTE (CD\_UTENTE) | Rappresenta la relazione fra EXPREG e UTENTE |
| EXPPLUS | VARCHAR |  | L'espressione regolare vera e propria, di tipo PLUS |
| EXPMINUS | VARCHAR |  | L'espressione regolare vera e propria, di tipo MINUS |

Tabella 3 - EXPREG

###### Le Espressioni Regolari esprimono il fatto che l'utente è o meno interessato in un certo argomento: le espressioni di TIPO PLUS indicano che, se nel bando compare una porzione di testo che fa match con tutte, bisogna notificare l'utente. Quelle di tipo MINUS indicano invece che NON bisogna notificare l'utente se se ne trova almeno una che fa match. La combinazione dei due tipi costituisce l'"argomento" di interesse.

## UTENTE

###### L'entità utente rappresenta gli account che sono iscritti al sistema e che desiderano ricevere delle notifiche in seguito alla pubblicazione di alcuni tipi di bando.

|  |  |  |  |
| --- | --- | --- | --- |
| Nome Proprietà | Tipo | Null / Extra | Commenti |
| CD\_UTENTE | BIGINT | PRIMARY, AUTO\_INCREMENT | Identificativo dell'utente, assegnato in automatico. |
| USERNAME | VARCHAR (50) | UNIQUE | Pseudonimo (univoco) dell'utente, scelto dallo stesso in fase di registrazione |
| PASSWORD | VARCHAR (255) |  | Password decisa dall'utente in fase di registrazione e di cui si calcola la funzione Hash MD5 |
| EMAIL | VARCHAR |  | Email decisa in registrazione |
| DT\_NOTIFICA | DATETIME |  | Data in cui è avvenuta l'ultima notifica di nuovo bando |

Tabella 4 - Entità UTENTE

# Process.pngProcessi

###### Le fasi del processo necessarie per raggiungere l'obiettivo sono almeno 4. Le varie fasi comunicano fra loro tramite dei messaggi e impostando uno stato, cioè proprietà delle entità Pubblicazione e Bando; una fase si "attiva" quando quella precedente termina.

## RICERCA NUOVA PUBBLICAZIONE

###### Una volta al giorno si collega al sito web della Gazzetta Ufficiale e cerca se è presente una nuova pubblicazione, controllandone la data e confrontandola con quella dell'ultima pubblicazione memorizzata dal sistema.

### In caso ne sia presente una nuova, crea una nuova entry nell'entità Pubblicazione e memorizza le prime informazioni già disponibili (numero, data e URL) . Il processo può così avanzare; nel caso non sia presente nessuna nuova pubblicazione, non sono attivate le fasi successive del processo.

## SCARICA PUBBLICAZIONE

###### Essendo disponibile una nuova pubblicazione, scarica la pagina corrispondente e analizza la pubblicazione, ricavando tutti i bandi e inserendoli nel database insieme a URL corrispondente, nome dell'ente richiedente e testo per intero. Queste informazioni sono già disponbili all'interno della pubblicazione.

### Una volta disponibili tutti i link ai bandi, sono scaricati uno per uno, aggiornando con nuove informazioni la precedente entry creata. La Pubblicazione può ora essere rimossa, in quanto tutti i bandi che conteneva sono stati scaricati.

###### Questo meccanismo permette di avere un controllo sull'effetivo scaricamento di tutti i bandi della pubblicazione: se tutti i suoi bandi sono stati scaricati, allora la pubblicazione non dovrebbe più essere presente nel database; se invece è presente anche solo un bando di una pubblicazione e la pubblicazione stessa, con relativo stato SCARICATA, significa che c'è stato un malfunzionamento e quindi sarebbe opportuno ripetere tutta l'operazione di scaricamento.

## PARSIFICA BANDI

###### Una volta disponibile il testo del bando, si parsifica e analizza per cercare di ricavare più informazioni possibili: oggetto del bando, CIG e descrizione; si separano anche logicamente nella struttura di memorizzazione. Queste informazioni, se presenti e se ricavabili, sono memorizzate all'interno dell'entità Bando, pronte per la prossima fase del processo.

###### Se il Parser non è stato in grado di ricavare le informazioni dal bando, perchè non presenti o male espresse o perchè il documento non era strutturato in un modo riconoscibile, rimane il testo integrale che può comunque essere utilizzato per la fase sucessiva.

## NOTIFICA BANDI

###### Per ogni utente memorizzato la cui proprietà DT\_NOTIFICA risulta precedente o uguale a DT\_INSERIMENTO (bando), analizza i nuovi bandi disponibili e controlla se rispettano o no le preferenze espresse.

### Ad ogni utente sono associate diverse espressioni regolari, tutte in OR fra loro, che devono o non devono comparire nel documento. Per ognuna di queste si tenta il match con l'oggetto (o il testo del documento, se l'oggetto manca): nel caso di espressioni MINUS, il match di una sola di queste comporta la fine immediata del controllo: l'utente non è notificato. Se, a operazione conclusa, una o più stringhe contenute nel bando hanno fatto match con almeno un'espressione regolare PLUS e non con quelle MINUS, l'utente è notificato. In altre parole, tutte le espressioni PLUS in OR fra loro vengono messe in AND con le espressioni di tipo MINUS (in OR fra loro) e tutte insieme vengono processate per il match con il bando, con la differenza che il match con il secondo tipo comporta il "fallimento" del riconoscimento.

### In seguito alla notifica di un utente, aggiorna il valore della proprietà DT\_NOTIFICA (nell'entità utente) con la data e l'ora dell'invio della notifica.

###### Il controllo di data è necessario in quanto, se il sistema non è stato in grado di notificare l'utente nel giorno stesso in cui è stato aggiunto il bando, l'operazione può essere eseguita in un secondo momento.

###### Se nessun bando interessa agli utenti registrati (secondo i criteri appena descritti), questa fase si conclude; se invece è presente almeno un bando che rispetta i parametri di uno o più utenti, sono notificati sul sito web. Se l'utente ha scelto di ricevere le notifiche anche tramite e-mail, il sistema provvede ad informare l'utente anche secondo questa modalità

###### Terminata questa fase, la catena di processi è conclusa e si deve aspettare la disponibilità di una nuova pubblicazione prima che tutto riparta.

# Dettagli Realizzazione Tecnica

## ARCHITETTURA

###### Per la realizzazione del progetto software si prevede di strutturare due differenti applicazioni:

### la prima si occupa di raccogliere i dati relativi ai bandi, elaborarli e memorizzarli. La seconda svolge il ruolo di presentazione dei risultati ottenuti, cioè la notifica degli utenti interessati; gestisce anche tutto quello che riguarda gli utenti e la loro interazione col sistema, il sito web e la posta elettronica.

### L'applicazione **GarePubblicheCrawler** è composta da tre moduli che interagiscono tra loro e corrispondono circa alle prime tre fasi del processo sopra descritte: ricerca nuova pubblicazione, scarica nuova pubblicazione, parsifica bandi.

###### Oltre a questi componenti, che realizzano la prima parte del sistema, va considerata l'applicazione **GarePubblicheWeb**: questa applicazione web svolge tre ruoli principali:

###### Permette ai nuovi utenti di registrarsi, di impostare e modificare i parametri di preferenza dei bandi e verifica gli accessi da parte di quelli registrati

###### Gestisce l'intero sito web e l'interfaccia che permette agli utenti di entrare nella loro area privata e di visualizzare i bandi di loro interesse.

###### Pubblica i risultati appena elaborati: notifica un utente che è disponibile un nuovo bando, ma solo se questo rispetta i parametri di interesse definiti dall'utente stesso. Espone i risultati sul sito web in un'area dedicata a ciascun utente, ed eventualmente invia per mail le notifiche dei bandi agli utenti che lo hanno richiesto.

## GESTIONE STATI

###### Nella prima fase del processo, appena dopo il rilevamento una nuova pubblicazione e la creazione della nuova entry nella tabella pubblicazioni, il valore dello stato di questa pubblicazione è **DA\_SCARICARE**, a significare che è pronta per essere scaricata.

### Nella fase di scaricamento si prendono in considerazione soltanto le pubblicazioni il cui stato è **DA\_SCARICARE** (dovrebbe essere una sola). La pubblicazione è quindi scaricata e processata per ricavare i bandi che contiene. Alla fine di questa fase, quando tutti i bandi sono stati ricavati e la fase è conclusa, il valore dello stato diventa **SCARICATA** e può quindi essere rimossa.

###### Nella fase di scaricamento della pubblicazione, quando i bandi sono ricavati e scaricati dalla pubblicazione, si crea una nuova entry della tabella Bandi, con stato **DA\_PARSIFICARE**.

### Il Parser tratta soltanto i bandi il cui stato è **DA\_PARSIFICARE** e procede con la loro analisi, al termine della quale lo stato del bando diventa **PARSIFICATO**.

### Tutti i bandi **PARSIFICATO** sono pronti per essere pubblicati: l'ultima fase prende i bandi con stato **PARSIFICATO** e li analizza per decidere a quali utenti notificare i nuovi bandi.

## STACK TECNOLOGICO

###### Gli strumenti tecnologici a disposizione sono: Java J2EE, Maven, Spring, Spring MVC, Prime Faces e MySQL.

###### Per la prima parte dell'applicazione, **GarePubblicheCrawler**, non vi è alcuna parte di presentazione, pertanto è utilzzato Java, accompagnato da Maven per la gestione delle dipendenze, MySQL per la memorizzazione dei dati e eventuali librerie esterne di utilità.

###### La seconda parte, **GarePubblicheWeb**, è invece un'applicazione web che si occupa della gestione di un sito; è quindi necessario uno stack più ampio:

###### Java J2EE, (Maven), MySQL, Spring e Spring MVC per quanto riguarda la parte di back end, di gestione delle pagine dinamiche e del database

### Prime Faces come framework per il front end