**NOME PROGETTO**

SOTTOTITOLO

**Obiettivo e ambito del documento**

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

**Documenti di riferimento**

| **Nome** | **Documento** |
| --- | --- |
| Esempio di pubblicazione della Gazzetta Ufficiale | [Esempio pubblicazione.htm](resources/Esempio%20pubblicazione.htm) |
| Esempio di bando nella pubblicazione | [Esempio Bando.htm](resources/Esempio%20Bando.htm) |

**Premessa**

L'obiettivo del progetto è di raccogliere e memorizzare i **bandi delle gare pubbliche** sulla **Gazzetta Ufficiale**, per poi notificare agli utenti registrati al sistema la presenza di nuovi bandi appena questi sono disponibili; gli utenti potranno esprimere delle preferenze sui tipi di bando a cui sono interessati e venire notificati soltanto se i bandi rispecchiano i loro parametri.

Le **Entità di Business** coinvolte dal progetto sono (la Gazzetta Ufficiale ,) le pubblicazioni della Gazzetta e i bandi pubblici contenuti e il sito web della Gazzetta Ufficiale, nel quale reperire le informazioni riguardanti i bandi. Di seguito descritte nei dettagli.

La Gazzetta è un documento ufficiale pubblicato dalla Republica Italiana con cadenza al massimo giornaliera. Le sue pubblicazioni, identificate da data e un numero (generalmente fra 1 e 153, si azzera con la fine dell'anno), espongono, oltre a decreti, sentenze e regolamenti, diverse decine di bandi di vario tipo, offerti da enti anche molto diversi quali Regioni, Province, Comuni ma anche Ministeri, Autorità indipendenti e altre figure soggettive pubbliche e private.

Le pubblicazioni dei bandi pubblici della Gazzetta (degli ultimi 30 giorni) vengono esposte nel sito web della Gazzetta Ufficiale, al seguente URL: "<http://www.gazzettaufficiale.it/30giorni/contratti>".  
Qua è possibile scegliere una pubblicazione e consultarne la pagina, che contiene l'elenco dei bandi e un collegamento agli stessi.

Da questa pagina si può inoltre accedere all'archivio completo di tutte le pubblicazioni.

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 interessano dei bandi sono sorattutto il **CIG** **(Codice Identificativo Gara)** e **l'oggetto della gara**.  
  
  
  
Analizzando i bandi si nota che purtroppo non rispettano una struttura uniforme e definita. però si possono classificare in due o tre "categorie" (forme in cui vengono redatti):

* La prima categoria 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, pertanto è facile trovare l'oggetto del bando, il suo codice e altre informazioni utili.  
  Esempio:  
    
  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 categoria strutturata, ma suddivisa soltanto in sezioni (in cui può o meno comparire il titolo della sezione). Anche in questa categoria di documenti risulta comunque non troppo complesso ricercare l'oggetto dell'appalto e il CIG (se è presente!)

Esempio:

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

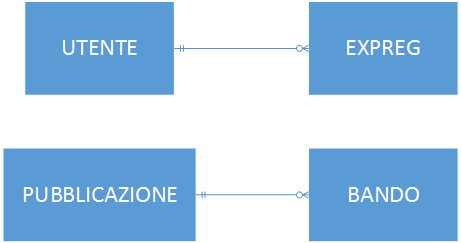
SEZIONE II: Oggetto

SEZIONE II:

* L'ultima categoria 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.

Una volta analizzati questi documenti si vuole, come già detto, estrarre il CIG e l'oggetto del bando, insieme a quante più informazioni possibili (come per esempio il nome dell'ente richiedente), per poi memorizzarli in modo persistente.   
  
Il sistema prevederà un meccanismo di registrazione dei nuovi **utenti**, con il loro username, password e la loro area privata, e la definzione da parte loro di alcuni parametri che rappresentano le loro **preferenze** riguardo a certi tipi di bando.  
Gli utenti possono definire quali sono gli "argomenti" che interessano, cioè delle espressioni da ricercare nel testo del bando (espressioni di tipo PLUS), e quali invece sono gli argomenti trattati dai bandi di cui non si vuole essere notificati (espressioni di tipo MINUS), in modo che, se nel bando compaiono determinate espressioni, l'utente viene subito scartato dalla lista dei potenziali interessati.  
Quando vengono pubblicati sulla Gazzetta dei nuovi bandi, dopo essere stati analizati e memorizzati, il sistema sceglierà quali sono gli utenti registrati interessati ai nuovi bandi (analizzando il testo del bando e cercando se sono presenti o no le "parole chiave" definite dagli utenti) e li notificherà.  
Le **notifiche** avverranno tramite servizi Internet (posta elettronica o un sito web), pertanto in fase di registrazione l'utente dovrà definire anche le modalità preferite di ricezione delle notifiche.

**Modello Entità-Relazione**



**PUBBLICAZIONE**

Permette di modellare le pubblicazioni della Gazzetta Ufficiale

|  |  |  |  |
| --- | --- | --- | --- |
| Nome Proprietà | Tipo | Null / Extra | Commenti |
| CD\_PUBBLICAZIONE | NUMBER | PRIMARY | Identificativo della pubblicazione. Viene popolato automaticamente |
| DATA | DATETIME |  | Data dell'inserimento della pubblicazione sul sito |
| NUMERO\_PUBBLICAZIONE | NUMBER (3) |  | Numero della pubblicazione |
| STATO\_PUBBLICAZIONE | VARCHAR (15) |  | Stato interno di gestione della pubblicazione ("Da Scaricare", "In Corso", "Scaricata" o "Chiusa") |
| URL | UriType |  | 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 | SEQUENCE | PRIMARY | Identificativo del bando, generato in automatico |
| CODICE\_ESTERNO | CHAR (11) | UNIQUE | Codice del bando che si trova nella pubblicazione (TX17BFC9332) |
| CD\_PUBBLICAZIONE | NUMBER | REFERENCES PUBBLICAZIONE (CD\_PUBBLICAZIONE) | Rappresenta la relazione fra BANDO e PUBBLICAZIONE |
| CIG | CHAR (10) | Null | Codice Identificativo Gara, riportato sul bando |
| TIPO\_BANDO | VARCHAR (50) | Null | Bando di Gara / Estratto di bando di gara / Bando di gara a procedura aperta |
| TIPO\_RICHIEDENTE | VARCHAR (100) | Null | Ministeri / Regioni / Province / ... |
| NOME\_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 | VARCHAR | Null | Tutto il testo integrale del bando |
| URL | UriType |  | Indirizzo web identificativo del bando |
| STATO\_BANDO | VARCHAR (20) | ENUM | Stato interno per la gestione ("Da Parsificare", In Corso", "Parsificato" o "Chiuso") |

**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 sia univoco.

Lo STATO è necessario per questioni tecniche del processo: i bandi appena scaricati vengono marcati come "da parsificare"; quando il Parser li avrà processati imposterà il valore dello stato a "parsificati" e quindi saranno pronti per la pubblicazione; infine quando verrà completata con successo la loro pubblicazione, il valore della proprietà diventerà "chiuso".

**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 | NUMBER | PRIMARY | Identificativo dell'espressione, assegnato in automatico. |
| CD\_UTENTE | NUMBER | REFERENCES UTENTE (CD\_UTENTE) | Rappresenta la relazione fra EXPREG e UTENTE |
| ESPRESSIONE | VARCHAR |  | L'espressione regolare vera e propria |
| TIPO\_ESPRESSIONE | VARCHAR (5) |  | PLUS / MINUS : Indica se il match comporta l'esclusione dell'utente o la sua eventuale notifica |

**Tabella 3 - EXPREG**

Le Espressioni Regolari esprimono il fatto che l'utente sia interessato o meno in un certo argomento: per le espressioni di TIPO PLUS indicano che, se nel bando compare una porzione di testo che fa match con questa, bisogna notificare l'utente. Quelle di tipo MINUS indicano ivece che NON bisogna notificare l'utente e si trova qualcosa 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 desierano ricevere delle notifiche in seguito alla pubblicazione di alcuni tipi di bando.

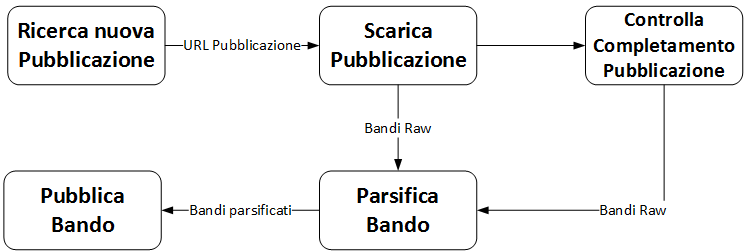
|  |  |  |  |
| --- | --- | --- | --- |
| Nome Proprietà | Tipo | Null / Extra | Commenti |
| CD\_UTENTE | NUMBER | PRIMARY | Identificativo dell'utente, assegnato in automatico. |
| USERNAME | VARCHAR (50) | UNIQUE | Pseudonimo (univoco) dell'utente, scelto dallo stesso in fase di registrazione |
| PASSWORD | VARCHAR (50) |  | Password decisa dall'utente in fase di registrazione |
| EMAIL | VARCHAR |  | Email decisa in registrazione |

**Tabella 4 - Entità UTENTE**

Necessaria una fase di registrazione utente in cui vengono inseriti tutti i dati necessari per le proprietà.

**Processi**

I processi necessari per raggiungere l'obiettivo richiesto sono almeno 5. Questi comunicano fra loro tramite dei messaggi e/o impostando uno stato (proprietà delle entità Pubblicazione e Bando) e sono indipendenti: si "attivano" quando le entità assumono un certo stato.



* **Ricerca Nuova Pubblicazione**:  
    
   Un "processo a parte", sempre in esecuzione: una volta al giorno si collega al sito web della Gazzetta Ufficiale (all'URL descritto nella premessa) e cerca se è presente una nuova pubblicazione, controllando la data della pubblicazione e confrontandola con quella dell'ultima memorizzata dal sistema.  
   In caso ne sia presente una nuova, crea una nuova entry di pubblicazione e memorizza le prime informazioni, già disponibili (numero, data e URL) e poi "notifica" il prossimo componente in modo che il processo avanzi; nel caso contrario non parte il resto del processo, ma si aspetterà fino a quando la ricerca non darà un risultato positivo.
* **Scarica Pubblicazione**:Questa fase del processo può essere raggiunta soltanto quando la ricerca di una nuova pubblicazione ha dato esito positivo. Essendo disponibile una nuova pubblicazione, viene scaricata la pagina identificata dall'URL che ha ricavato il passo precedente del processo, poi viene analizzata la pubblicazione, ricavati e scaricati tutti i bandi, ognuno con il proprio URL, nome dell'ente richiedente (già disponbile all'interno della pubblicazione) e testo per intero (Bando Raw nello schema sopra).

* **Controlla Completamento Pubblicazione**:  
    
  Siccome non ci sono garanzie sul completamento dei processi e potrebbe capitare un qualsiasi evento che potrebbe compromettere un processo e la sua riuscita, viene controllato che la pubblicazione sia effettivamente stata scaricata e che i bandi che conteneva siano stati memorizzati tutti.   
  Per questo scopo è necessario, appena la pubblicazione è disponibile e in seguito al completamento (presunto) della fase precedente, effettuare nuovamente l'operazione di scaricamento della pubblicazione, per poi controllare che il numero dei bandi che contiene corrisponda al numero di quelli memorizzati corrispondenti all'ultima pubblicazione.  
  In caso di esito negativo vengono rimossi tutti gli ultimi bandi salvati e viene ripetuto il processo di scarciamento dei bandi e di ricavo delle loro informazioni preliminari.  
  Altrimenti questa fase termina e si passa alla parsificazione dei bandi.
* **Parsifica Bando**  
    
  Una volta disponibile il testo del bando, viene parsificato e analizzato per cercare di ricavare più informazioni possibili: oggetto del bando, CIG e descrizione; vengono anche separate logicamente nella struttura di memorizzazione. Queste informazioni (se presenti e se ricavabili) vengono memorizzate all'interno dell'entità Bando, pronte per la prossima fase del processo. Se il Parser non è stato in grado di ricavare queste 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 potrà comunque essere utilizzato poi.

* **Pubblica Bando**Analizza l'oggetto del bando (e eventualmente anche il testo) e controlla se rispetta o no le preferenze espresse dagli utenti, e quali. Ad ogni utente sono associate diverse espressioni che devono o non devono comparire nel documento: viene eseguita una ricerca per determinare se il documento fa match con le espressioni.  
  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, questi vengono notificati: a seconda che l'utente abbia scelto di ricevere le notifiche via mail o tramite il sito web dell'applicazione, questa fase informa l'utente secondo la modalità da lui scelta.  
  Terminata questa fase la catena di processi è conclusa e si dovrà aspettare nuovamente la disponibilità di una nuova pubblicazione (prima fase) prima che tutto riparta.

**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 vengono modificati nelle diverse fasi dei processi ed esprimono un'informazione utile alle altre fasi, che si comporteranno in modo diverso a seconda del valore dello stato.  
Rappresentano quindi il modo in cui le fasi comunicano fra loro, scambiandosi un messaggio.

**PUBBLICAZIONE**

|  |  |  |
| --- | --- | --- |
| Nome Stato | Valore | Commenti |
| STATO\_PUBBLICAZIONE | "Da scaricare" | Pubblicazione pronta da scaricare |
| "In Corso" | La pubblicazione è in fase di scaricamento / processamento |
| "Scaricata" | La pubblicazione è stata scaricata e i suoi bandi pronti ad essere parsificati |
| "Chiusa" | Si è concluso il processamento della pubblicazione e tutti i bandi che conteneva sono stati ricavati e memorizzati |

Nella prima fase del processo, appena viene rilevata una nuova pubblicazione e creata la nuova entry nella tabella pubblicazioni, il valore dello stato della pubblicazione è impostato a "**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). Si inizia impostando lo stato della pubblicazione a "**In corso**": viene poi scaricata e processata per ricavare i bandi. Alla fine di questa fase, quando tutti i bandi sono stati ricavati e la fase è conclusa, il valore dello stato diventa "**Scaricata**".  
  
La fase di controllo della pubblicazione (appena successiva alla fase di scaricamento), prenderà in considerazione soltanto l'ultima pubblicazione inserita (con data più recente) il cui stato è "**Scaricata**".

Successivamente, quando finisce la fase di parsificazione dei bandi, lo stato della pubblicazione viene impostato a "**Chiusa**", a significare che tutti i bandi della pubblicazione sono stati processati e che quindi non è più necessario operare in nessun modo sulla pubblicazione.

**BANDO**

|  |  |  |
| --- | --- | --- |
| Nome Stato | Valore | Commenti |
| STATO\_BANDO | "Da Parsificare" | Il bando è stato scaricato tutto ed è pronto per essere parsificato |
| "In Corso" | Il bando è in fase di analisi da parte del Parser |
| "Parsificato" | Si è conclusa l'analisi del bando ed è pronto per essere pubblicato |
| "Chiuso" | Tutti gli utenti potenzialmente interessati al bando sono stati notificati |

Nella fase di scaricamento della pubblicazione (e nella fase di controllo di completamento della pubblicazione), quando i bandi vengono ricavati e scaricati dalla pubblicazione, viene creata una nuova entry della tabella Bandi e lo stato di questi è "**Da Parsificare**".  
Il Parser tratterà soltanto i bandi il cui stato è "**Da Parsificare**" e subito setterà il loro stato a "**In Corso**"; poi procederà con la loro analisi, al termine della quale lo stato del bando diventerà "**Parsificato**".  
Tutti i bandi "**Parsificato**" sono pronti per essere pubblicati: l'ultima fase cercherà nei bandi con stato "**Parsificato**" e, dopo averli analizzati e deciso a quali utenti notificare i nuovo bandi, imposterà lo stato a "**Chiuso**": non verranno più trattati in nessuna fase.

**Dettagli Realizzazione Tecnica**

Per la realizzazione del progetto software si prevede di strutturare due differenti applicazioni:  
la prima si occupa di raccogliere i dati relativi ai bandi e di memorizzarli; la seconda svolge il ruolo di analisi dei bandi e di notifica degli utenti interessati, gestendo anche tutto quello che riguarda gli utenti e la loro interazione col sistema.  
I componenti delle diverse applicazioni corrispondono all'incirca ai processi sopra descritti (un componente per processo).

L'applicazione di **RaccoltaGarePubbliche** è composta da 5 moduli che interagiscono tra loro:

* Un componente che, una volta al giorno (indicativamente), controlla il sito web della Gazzetta Ufficiale e cerca se sono state inserite nuove pubblicazioni; in caso di esito positivo ne memorizza l'URL.
* Un modulo che, quando è disponibile una nuova pubblicazione, la scarica, analizzandola ed estraendone le proprietà, per poi memorizzare anclune informazioni preliminari sui bandi che contiene (per esempio il loro indirizzo e il nome dell'ente richiedente, insieme al testo integrale del bando).
* Per assicurarsi che siano stati scaricati correttamente tutti i bandi, un componente controlla che siano effettivamente presenti nel sistema tutti i bandi di una certa pubblicazione; in caso contrario verrà ripetuta l'operazione precedente di scaricamento di tutti i bandi.
* Il Parser, che si occupa di leggere i bandi, ricavare le informazioni di interesse e strutturarle secondo un certo schema all'interno del sistema di memorizzazione.

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

* Pubblica i risultati appena elaborati: notifica un utente (tramite sito web o posta, a seconda della preferenza dell'utente) che è disponibile un nuovo bando, ma solo se questo rispetta dei parametri di interesse definiti dall'utente stesso.
* Gestisce l'interfaccia che permette agli utenti di entrare nella loro area privata e di visualizzare i bandi di loro interesse.
* 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.