# 1 Introduzione

## Scopo Requirement Document

Lo scopo di questo documento è quello di fornire una documentazione il più dettagliata possibile riguardo le specifiche dei requisiti del progetto *A* *Recommender System for News* (RSN), sistema che fornisce una lista di news all’utente in base alle sue preferenze; le news proposte all’utente sono tutte prese dai tweet degli account Twitter di importanti testate giornalistiche. L’obiettivo finale è quello di realizzare una web app a cui chiunque possa registrarsi ed accedere per essere aggiornato sulle notizie riguardanti gli ambiti a cui è più interessato.

In questo documento verranno illustrate le features e altre informazioni riguardanti il progetto in modo da fornire una visione d’insieme del prodotto che vogliamo sviluppare.

## Scopo del progetto

L’idea che sta alla base del progetto è quella di creare un’applicazione disponibile sul web a qualsiasi utente sia interessato a notizie giornalistiche che possono coprire diversi ambiti; lo scopo principale consiste nel mettere in evidenza notizie di determinate categorie (es. sport, politica, etc.) piuttosto che altre in base agli interessi dei diversi utenti. Per poter eseguire questa operazione è necessaria una fase di profilazione degli utenti che avverrà durante la registrazione all’applicazione.

Il progetto può essere suddiviso in due macroaree:

* un crawler: questa parte riguarda la raccolta dei dati, ossia le news, dal social network Twitter che utilizzeremo come “database”;
* un algoritmo per l’analisi, indicizzazione e raccomandazione delle news.

## Definizioni, acronimi, abbreviazioni

RSN – A Recommender system for news – nome del progetto/applicazione

## Riferimenti

## 1.5 Overview

# 2 Descrizione generale

## 2.1 Visione del prodotto

### 2.1.1 Ambito del progetto

L’applicazione (RSN) si colloca nell’ambito di reperimento informazioni legate a news reperite da twitter e può renderle visibili all’utente in base agli interessi di quest’ultimo. Il progetto è suddiviso in due parti: crawler delle news e metadati, e analisi tramite algoritmo di raccomandazione dei dati raccolti.

### 2.1.2 Ambiente

L’applicazione finale funzionerebbe in prima battuta su qualsiasi tipo di browser cosicché l’utente sia in grado di autenticarsi sul nostro sistema ovunque e con qualsiasi dispositivo.  
I dati che forniremo all’utente sono provenienti dagli account Twitter di alcune delle più note testate giornalistiche.

L’ambito di interesse è Twitter in quanto lo scopo del progetto è rendere disponibili i dati raccolti dal social network e a questo scopo sono necessarie le api-twitter per poter creare un crawler per raccogliere i dati.

L’applicazione finale funzionerebbe in prima battuta su qualsiasi browser. Cosicché l’utente abbia la possibilità di loggarsi ovunque e da qualsiasi dispositivo.

### 2.1.3 Evoluzioni future

Il primo passo sarà quello di implementare una maggiore fonte di notizie andando a raccogliere dati da fonti differenti da Twitter.  
L’evoluzione futura di questa webApp prevede la creazione dell’applicazione su piattaforme mobile come Android o IOS.

### 2.1.4 Architettura

Crawler + algoritmo = dati raccomandati <- app

Server...

(fare schemino)

### 2.1.5 Confronto con business-concorrents (altri modelli simili)

Flipboard

## 2.2 Dati e funzioni del software

Abbiamo deciso di creare un elenco di testate giornalistiche da cui prelevare le informazioni. La scelta è stata basata su una ricerca per definire le fonti più autorevoli e più attive sul social network Twitter così da avere un aggiornamento delle informazioni quotidiano.

Per definire le più autorevoli abbiamo consultato il dato relativo alla diffusione sia cartacea che digitale della testata; affidandoci a questi dati siamo poi andati a verificare la frequenza di pubblicazione (attività) sul social network Twitter.

Le testate che abbiamo ritenuto soddisfacenti per le nostre esigenze sono:  
@Corriere 2,1Mln 326000

@sole24ore 1,34Mln 111000

@repubblica 2,84Mln 581000

@LaStampa 1,03Mln 358000

@fattoquotidiano 1,92Mln 138000

@ilmessaggeroit 390000 247000

@Avvenire\_Nei 62800 538007

| **Testata** | [**Tiratura**](https://it.wikipedia.org/wiki/Tiratura) | [**Diffusione cartacea**](https://it.wikipedia.org/wiki/Diffusione_(editoria)) | **Diffusione digitale** | **Totale diffusione** |
| --- | --- | --- | --- | --- |
| [Il Sole 24 ORE](https://it.wikipedia.org/wiki/Il_Sole_24_ORE)[[2]](https://it.wikipedia.org/wiki/Quotidiani_in_Italia#cite_note-old-2) | 257 305 | 200 155 | 115 366 | 315 521 |
| [Corriere della Sera](https://it.wikipedia.org/wiki/Corriere_della_Sera) | 322 826 | 226 453 | 70 794 | 297 247 |
| [Quotidiano Nazionale](https://it.wikipedia.org/wiki/Quotidiano_Nazionale) | 309 212 | 218 413 | 3 183 | 221 596 |
| [la Repubblica](https://it.wikipedia.org/wiki/La_Repubblica_(quotidiano)) | 292 365 | 190 260 | 30 160 | 220 420 |
| [La Gazzetta dello Sport](https://it.wikipedia.org/wiki/La_Gazzetta_dello_Sport) | 252 250 | 166 998 | 10 659 | 177 657 |
| [La Stampa](https://it.wikipedia.org/wiki/La_Stampa) | 214 691 | 145 734 | 25 046 | 170 780 |
| [Avvenire](https://it.wikipedia.org/wiki/Avvenire) | 134 916 | 102 007 | 8 848 | 110 855 |
| [Il Messaggero](https://it.wikipedia.org/wiki/Il_Messaggero) | 144 337 | 100 569 | 8 477 | 109 046 |
| [QN - il Resto del Carlino](https://it.wikipedia.org/wiki/Il_Resto_del_Carlino) | 133 113 | 98 813 | 1 754 | 100 567 |
| [Corriere dello Sport-Stadio](https://it.wikipedia.org/wiki/Corriere_dello_Sport-Stadio) | 197 623 | 89 287 | 1 224 | 90 511 |
| [Italia Oggi](https://it.wikipedia.org/wiki/Italia_Oggi)[[2]](https://it.wikipedia.org/wiki/Quotidiani_in_Italia#cite_note-old-2) | 98 366 | 54 166 | 18 157 | 72 323 |
| [il Giornale](https://it.wikipedia.org/wiki/Il_Giornale) | 120 738 | 59 698 | 1 412 | 61 110 |
| [Tuttosport](https://it.wikipedia.org/wiki/Tuttosport) | 134 884 | 55 550 | 1 191 | 56 741 |
| [il Fatto Quotidiano](https://it.wikipedia.org/wiki/Il_Fatto_Quotidiano) | 85 441 | 35 105 | 10 662 | 45 767 |
| [Il Mattino](https://it.wikipedia.org/wiki/Il_Mattino) | 54 526 | 36 717 | 2 896 | 39 613 |
| [Libero](https://it.wikipedia.org/wiki/Libero_(quotidiano)) | 79 550 | 26 167 | 869 | 27 036 |
| [La Verità](https://it.wikipedia.org/wiki/La_Verit%C3%A0_(quotidiano)) | 66 631 | 21 500 | 595 | 22 095 |

### 2.2.1 Modello di dominio

I dati che reperiamo sono quelli rilevanti sia i singoli utenti che i singoli tweet:

* per gli utenti prevediamo di conservare i dati rilevanti l’autenticazione e le loro preferenze
* per i tweet i dati essenziali a dei possibili ordinamenti

Il modello di base è il seguente:

modello E/R del database di base.

### 2.2.2 Obiettivi funzionali di base

Download dei dati e metadati

Indicizzazione

Profilazione utente

Ordinamento per interessi derivati dalla profilazione

Log in

Register

### 2.2.3 Informazioni da prelevare

ID

Autore

Testo

Data creazione

favouriteCount

retweetCount

### 2.2.4 Memorizzazione dei dati

Le informazioni raccolte saranno memorizzate in una formattazione che permetta una gestione ottimale dei dati e metadati.

Xml……

### 2.2.5 Aggiornamento data base

Aggiunte di testate giornalistiche?

Ogni quanto aggiornare database?

Quando una news diventa obsoleta?

Se news obsoleta eliminazione da DB o non mostro all’utente?

## 2.3 Utenti e scenari di utilizzo

Utenti registrati (con profilazione)

Utenti non registrati (senza profilazione)

Utenti amministratori

Modello di dominio: tutte entità espliciti

Funzionalità, vincoli, entità

Modello dei dati

# 3 Requisiti specifici

## 3.1 Requisiti funzionali

### 3.1.1 Requisiti dell’interfaccia

L’applicazione avrà un’interfaccia grafica che permetterà all’utente di effettuare le seguenti azioni:

La nostra applicazione deve permettere ad un utente di registrarsi per la prima volta utilizzando semplice form di mail ecc. o tramite servizi facebook, gmail ecc.

L’app deve permettere una prima scelta delle preferenze ad un utente appena iscritto subito dopo la procedura di registrazione.

Deve permettere l’autenticazione di un utente già registrato mantenendo le preferenze di esso.

Deve permettere l’aggiornamento delle preferenze di ogni utente in base agli articoli che legge/apre.

Il sistema deve essere in grado di mostrare all’utente le notizie più rilevanti ai suoi interessi.

Ordinamento per interessi derivati dalla profilazione come default.

Possibilità di switchare tra ordinamento news in base a: data, preferenze utenti, numero like post, retweet, commentate.

Possibilità di aggiornare le notizie al momento.

Possibilità di logout.

### 3.1.2 Requisiti del server

Crawling tramite api-twitter

Indicizzazione dati news

Algoritmo di raccomandazione

Aggiornamento automatico news

## 3.2 Requisiti non funzionali