**PROGETTO OPEN DATA AND WEB SERVICES**

**Parte 1 – Open Data**

**Punto 1 – Qual è il problema che ci siamo posti?**

Il problema centrale è l’incertezza economica imprenditoriale.

Più informalmente, ci siamo messi nei panni di un imprenditore che voglia investire parte del proprio capitale.

**Punto 1.2 – il bisogno?**

Necessità di orientamento per le dinamiche di mercato.

**Punto 1.3 – gli open data posso rispondere a questo bisogno?**

Si, vedremo che esistono banche dati in grado di fornisci informazioni e dati relativi a tale bisogno.

**Punto 2 – Obiettivo generale**

Creare uno strumento che possa rilevare l’attrattività e la competitività tra i paesi europei, analizzandone e confrontandone dati economici di natura generale e settoriale.

**Punto 3– Obiettivo specifico**

Più nel dettaglio, di fornire un servizio che sia in grado di coadiuvare la scelta imprenditoriale da parte di uno stakeholder ovvero portatore di interessi (portatore di capitale).

Al fine di produrre un aiuto concreto per determinare la scelta ottimale, dapprima abbiamo definito una serie di parametri econometrici che definiscano il livello di “salute economica” all’interno dei paesi nell’Unione Europea negli ultimi cinque anni e, in seconda fase, abbiamo scelto alcuni possibili settori attraverso determinati indici in grado di generare un quadro di appetibilità per un possibile investimento.

**Punto 4 – Risultati attesi:**

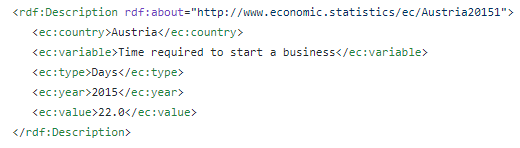
Non solo l’imprenditore riuscirà ad orientarsi al meglio nella scelta di un settore in un determinato paese europeo dove investire, ma potrà anche esaminare lo sviluppo di quel determinato indice nel corso dei cinque anni all’interno di uno stesso paese.

**Punto 5– Attività: quali passaggi abbiamo effettuato?**

1. Fase 1: **Ideazione del funzionamento del web service: come il web service aiuta l’imprenditore?**il web service sarà in grado di posizionare in una classifica i vari stati europei a seconda dei parametri economici generali per andare poi ad approfondire il livello di competitività in dei mercati specifici.

La chiave del funzionamento dell’intero web service è quella di poter comparare le nazioni europee a seconda dei mercati o dei parametri di mercato, creando una classifica per ogni esigenza. L’imprenditore sarà così alla fine dotato di una piattaforma di facile consultazione e con una dettagliata quantità di dati “raccolti anche in maniera cronologica” atta a renderlo più cosciente delle scelte che potrà effettuare.

1. Fase 2: **Identificazione dei parametri**: Il primo obiettivo è stato quello di identificare i parametri che fornissero un quadro economico generale del paese europeo. A tal proposito abbiamo scelto i seguenti indici:
   1. time to require to start a business;
   2. time to prepare and pay taxes;
   3. total tax and contribution rate;
   4. cost of business start-up procedures;
   5. exports of goods and services;
   6. import of goods and services;
   7. GDP per capita.
2. Più nel dettaglio,per ogni paese europeo abbiamo poi deciso di analizzare i seguenti **macrosettori** basandoci sul livello di fatturato italiano maggiore negli ultimi anni:
3. agricoltura;
4. automotive;
5. ricerca e sviluppo;
6. sostenibilità;
7. trasporti;
8. tecnologia.
9. Per ogni settore abbiamo analizzato i **principali parametri**:
   1. Agricoltura:
      1. Total agricultural land area;
      2. Area under organic farm;
      3. Output of agricultural industry.
   2. Automotive:
      1. Exports of consumption goods, motor spirit and passenger motor cars;
      2. Imports of consumption goods, motor spirit and passenger motor cars;
      3. Enterprises number of manufacture of motor vehicles, trailers and semi-trailers;
      4. World motor vehicle production;
      5. New registrations of passenger cars;
      6. Registrations or sales of new vehicles.
   3. Ricerca e sviluppo:
      1. Spesa interna lorda in R&S per settore di prestazione e obiettivo socioeconomico;
      2. Spese aziendali per R&S;
      3. Spese pubbliche di R&S finanziate dall’industria;
      4. Stanziamenti del bilancio pubblico per ricerca e sviluppo;
      5. Totale personale R&S Innovazione regionale.
   4. Sostenibilità:
      1. Final energy consumption;
      2. Greenhouse gas emissions;
      3. Exposure to air pollution by particulate matter;
      4. Eco-innovation index;
      5. Share of renewable energy in gross final energy consumption by sector;
      6. Recycling rate of municipal waste;
      7. Circular material use rate.
   5. Trasporti:
      1. Length of motorways and e-roads;
      2. Length of railway lines;
      3. Number of airports;
      4. Number of enterprises;
      5. Number of vessels in the main ports;
      6. Transport infrastructure investment spending;
      7. Transport infrastructure maintenance spending;
   6. Tecnologia:
      1. Imprese che impiegano specialisti in ICT;
      2. Quota del fatturato delle imprese sul commercio elettronico;
      3. Acquisti su internet da parte di privati;
      4. Aziende che danno dispositivi portatili per una connessione mobile a Internet ai propri dipendenti;
      5. Aziende con connessione a banda larga;
      6. Demografia aziendale per classe di dimensione;
      7. Livello di competenze digitali;
      8. Percentuale del settore ICT in PIL.
10. Fase 3: **Raccolta dei dati sul web**: Le banche dati dalle quali abbiamo estrapolato i dati necessari per la nostra analisi sono:
    1. OECD: <http://www.oecd.org/termsandconditions/> ;
    2. EUROSTAT: <https://ec.europa.eu/info/legal-notice_en> ;
    3. CIA: <https://www.cia.gov/about-cia/site-policies/> ;
    4. WEFORUM: <https://www.weforum.org/about/terms-of-use> ;
    5. FAOSTAT: <http://www.fao.org/contact-us/terms/db-terms-of-use/en/> ;
    6. THEWORLDBANK: <https://www.worldbank.org/en/about/legal> ;
    7. TRANSPARENCY INTERNATIONAL: <https://www.transparency.org/permissions> .
11. Fase 4: **Pulizia e organizzazione dei dati raccolti**: Dopo aver raccolto tutti i dati necessari, abbiamo iniziato la pulizia dei dati: abbiamo deciso di raccogliere in un unico file excel tutti gli indici per ogni settore. I dati si presentavano a primo acchito in modo corretto a pulito quindi non è stata necessaria una vera e propria pulizia di dati tramite ausili quali script Python etc…
12. Fase 5: **Creazione di file RDF:** Dopo aver raccolto i dati nei diversi file excel di ogni settore, abbiamo trasformato tali file .xls in file .rdf mediante l’ausilio del tool OpenRefine, la struttura dei file rdf è stata ideata come segue:



La 1 riga “<ec:country…. “ indica il paese preso in considerazione, nella 2 abbiamo deciso di inserire il nome del parametro, la 3 riga è composta dall’unità di misura di tale parametro, infine 4 e 5 riga contengono rispettivamente l’anno di riferimento e il valore numerico del parametro in oggetto.

**Parte 2 – Web Service**

**Fase 1: Creazione del servizio web (pom e web)**

La scelta per la costruzione del servizio web è ricaduta su un servizio di tipo REST; i software utilizzati nella creazione del Web service sono stati Eclipse (ambiente di sviluppo utilizzato), Apache CXF (framework di servizi Web), Apache Tomcat (server web open source) e Maven (strumento di build automation. Il linguaggio utilizzato è stato Java.

Il primo passo è stato dunque la creazione di un nuovo progetto Maven, EconomyRESTServiceMavenRDF, dopodiché si è proceduto alla creazione del file pom.xml, un file xml di configurazione che contiene tutte le informazioni sul progetto (dipendenze, test, documentazione, etc…).

In particolare un POM è composto da un tag root <project> che conterrà i tag:

* <modelVersion> che dichiara a quale versione di POM questo progetto è conforme;
* <groupId> ID del gruppo del progetto;
* <artifactId> ID dell’artefatto del progetto;
* <version> cioè la versione del progetto;
* <packaging> che è il tipo di archivio che vogliamo esportare (jar o war)

I parametri **groupId**, **artifactId**, **version** e **packaging** identificheranno univocamente un progetto.

Una volta creato il file pom.xml è richiesta la creazione del file web.xml (facilmente intuibile da un errore che si presenterà nella schermata del file pom su eclipse). Il file web.xml è necessario in quanto il nostro obiettivo è la creazione di un progetto web, quindi il nostro package è stato settato su war al momento del inizializzazione.

**Fase 2: Creazione del servizio web (Service)**

Dopo queste operazione preliminari abbiamo creato il nostro package al cui interno sono state costruite le interfacce e le classi di ogni settore.



All’interno di ogni interfaccia abbiamo richiamato i metodi che sono stati successivamente implementati nella relativa classe; i metodi da noi ideati sono stati i seguenti:

* Ricerca nell’RDF per Country;
* Ricerca nell’RDF per Year;
* Ricerca nell’RDF per Variable.

Queste sono le tre query più semplici con cui abbiamo deciso di iniziare la nostra investigazione degli RDF; in particolare la prima effettua una ricerca in ciascun RDF in base al paese che l’utente inserisce nella URL, la seconda effettua una ricerca per anno e la terza per nome della variabile presente in ciascun settore.

Proseguendo sulla scia delle indagini appena effettuate, abbiamo deciso di innalzare ulteriormente il livello di difficoltà delle nostre query per poterci addentrare ancor di più nell’analisi. Pertanto abbiamo deciso di implementare due nuovi metodi:

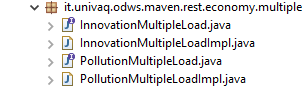
* Ricerca nell’RDF di paesi per Year;
* Ricerca nell’RDF di variabili per Country.

La prima query restituisce, all’inserimento da parte dell’utente di un determinato anno nella URL, un paese con l’elenco di tutte le sue variabili (chiaramente diverse da settore a settore) e i corrispettivi valori. La seconda query invece restituisce tutte le informazioni (anno, valore…) di una particolare variabile, scelta in fase di progettazione, dopo l’inserimento da parte dell’utente di un qualsiasi paese tra quelli presenti negli RDF (paesi UE).

Un ulteriore query che fornisce una ricerca interessante è un’altra tipologia di query introdotte nel progetto che effettuano una ricerca per anni, in particolare restituiscono tutte le informazioni e le variabili degli anni minori/maggiori dell’anno inserito dall’utente nella URL.

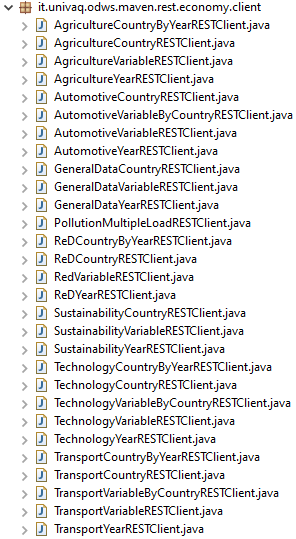
Infine l’ultima tipologia di query sperimentata è stata quella della combinazione di diversi dataset:

* PollutionMultipleLoad che mette in risalto lo stretto legame che vi è fra la produzione di veicoli a motore e l’esposizione all’inquinamento atmosferico;
* InnovationMultipleLoad NON SE SA.



**Fase 2: Creazione del servizio web (Client)**

Una volta testato il servizio e constatato che tutto funzionava correttamente siamo passati alla fase di creazione del client per il nostro progetto, EconomyRESTClientMavenRDF. Quindi anche qui abbiamo proceduto con la creazione del file pom.xml e dopodiché siamo passati alla creazione del nostro package dove abbiamo inserito tutte le classi.



Come è possibile vedere dalla figura sopra abbiamo rinominato ogni classe in base al metodo che essa implementa nel servizio, per ciascun settore. Ogni classe invoca in una stringa private l’url <http://localhsot:8080/EconomyRESTServiceMavenRDF/rest/> e in una stringa public l’url corrispettivo di ogni metodo; ad esempio per il metodo getAgricultureByCountry si avrà “agricolture/country/Italy”. In tal modo quando si farà il Run on Java Application del Client si vedrà nella console di Eclipse il risultato del metodo scelto al momento dell’invocazione.

**Alice Laraspata**

**Claudio Morelli**

**Simon Danny Pettinella**