

Modelarea/predicția prin serii temporale a cutremurelor în România

Raport final

Obiectivele proiectului

- Identificarea datelor statistice (INS) aferente seriei temporale
- Definirea componentelor de tip tendință/periodicitate/aleatorie din cadrul modelului
- Simularea și predicția seriei temporale pentru perioada 2017-2022

Privire de ansamblu asupra aplicației

Aplicația permite analiza și predicția evenimentelor seismice prin prisma seriilor temporale. Utilizatorul interacționează cu aceasta printr-o interfață grafică.

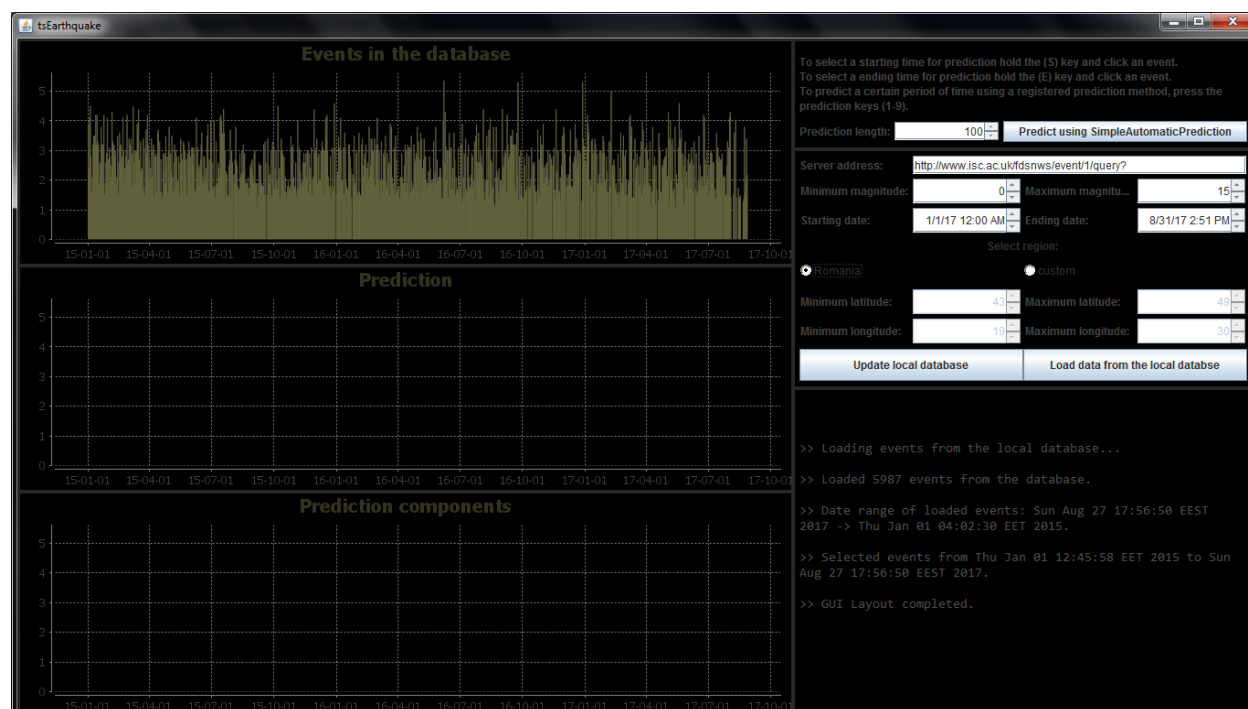
Modul de lucru este facil. Utilizatorul poate începe prin a descărca date de la un server dedicat de evenimente seismice în format QuakeML, sau poate lucra direct cu evenimentele din baza de date locală. Descărcarea de noi date nu implică cunoștințe avansate: utilizatorul introduce doar parametrii necesari evenimentelor și aplicația trimite o cerere HTTP către server, salvând automat răspunsul.

După stabilirea bazei de date de evenimente seismice, utilizatorul are opțiunea de a vizualiza seria temporală a acestora în interfața grafică. Pentru a genera predicții pe un anumit subset din seria temporală, utilizatorul selectează cu mouse-ul subsetul dorit, introduce, dacă este nevoie, parametrii predicției, și pornește predicția. La finalul predicției, aceștia îi vor fi prezentate rezultatele atât în graficele cu serii temporale cât și în consolă.

Interfața cu utilizatorul

Interfața grafică

Interfața grafică este ușor de folosit și are două zone principale: zona din stânga este rezervată graficelor seriilor temporale și evenimentelor seismice iar cea din dreapta este rezervată controlului aplicației.



Prima reprezentare grafică arată evenimentele disponibile în baza de date și evenimentele selectate spre a fi prelucrate. Acestea sunt afișate sub forma unui grafic cu bare, fiecare bară reprezentând evenimentul cel mai semnificativ din ziua respectivă. Cel de-al doilea grafic reprezintă într-o manieră asemănătoare evenimentele prezise. Al treilea grafic reprezintă componentele specifice metodei de predicție (serii temporale, tendințe, componente sezoniere, componente aleatorii, etc.). În momentul activării unei tip de predicție, acest grafic va primi o legendă a tuturor elementelor din el.

Graficele pot fi controlate prin intermediul mouse-ului. Prin acțiunea de tragere se poate schimba orizontul de timp al evenimentelor afișate și prin acțiunea de scroll se poate apropia sau îndepărta graficul.

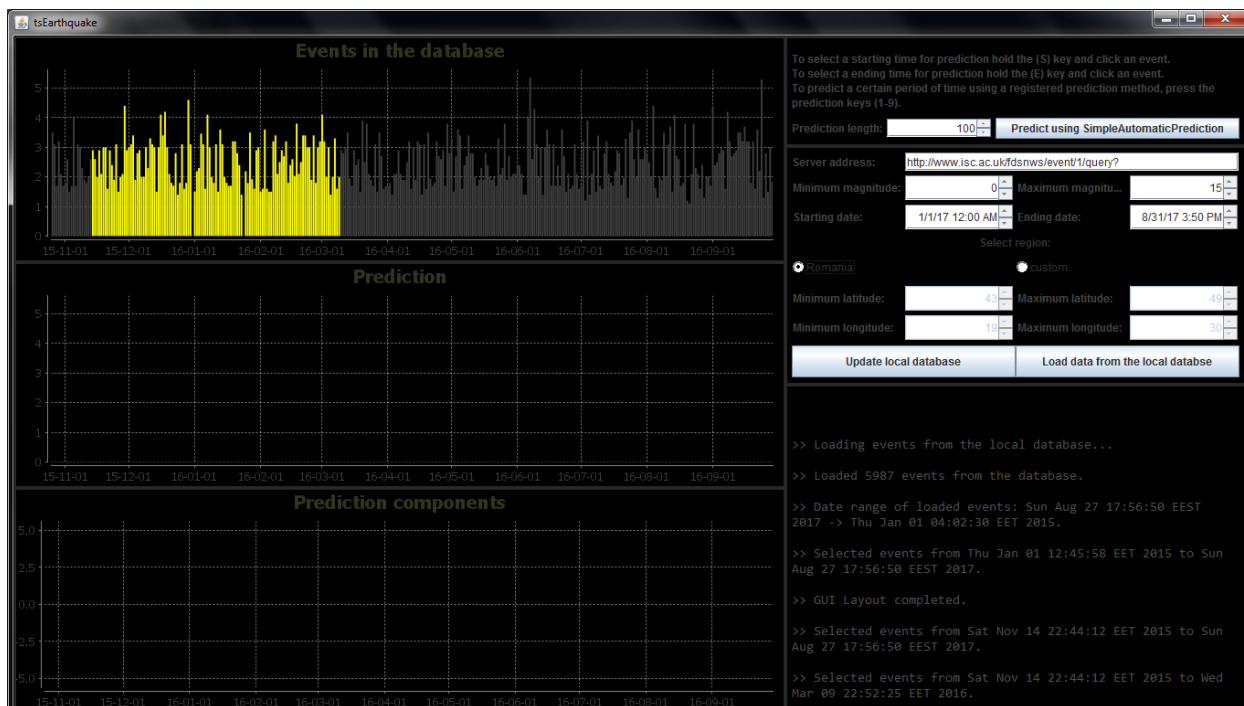
Partea desemnată controlului aplicației are trei zone. Zona principală prezintă informații de bază despre modul de utilizare al aplicației și parametri și controale specifice predicțiilor. Zona din mijloc este dedicată încărcării de evenimente în baza de date. Aceasta are parametri necesari accesării bazei de date din care se încarcă evenimentele. Zona din partea de jos este o consolă în care sunt afișate informații despre activitatea utilizatorului, evenimente ale aplicației și mesaje, informații și evenimente alte predicției.

Predicția seriei temporale

Utilizatorul generează o predicție a seriei temporale prin intermediul zonei principale de control. După introducerea parametrilor predicției (minim lungimea acesteia), utilizatorul apasă butonul de prezicere și așteaptă rezultatul acesteia. Rezultatul este încărcat în grafice și afișat automat, iar mesajele generate și informațiile necesare sunt afișate în consolă. Selectarea tipului de predicție înregistrată se face apăsând tastele de predicție (tastele 0-9).

Selectarea evenimentelor

Utilizatorul are opțiunea de a genera o predicție folosind numai un subset al datelor încărcate. Selectarea evenimentelor se face cu mouse-ul, din graficul principal de evenimente. Selecția se face alegând un eveniment de pornire și unul de final din setul de evenimente. Activarea selectării evenimentului de start al seriei se face prin menținerea apăsată a tastei S și activarea selectării evenimentului de final al seriei se face prin menținerea apăsată a tastei E. Selectarea evenimentelor se face apăsând pe acestea. Combinația de utilizare de mouse și taste crește ergonomia interfeței programului.



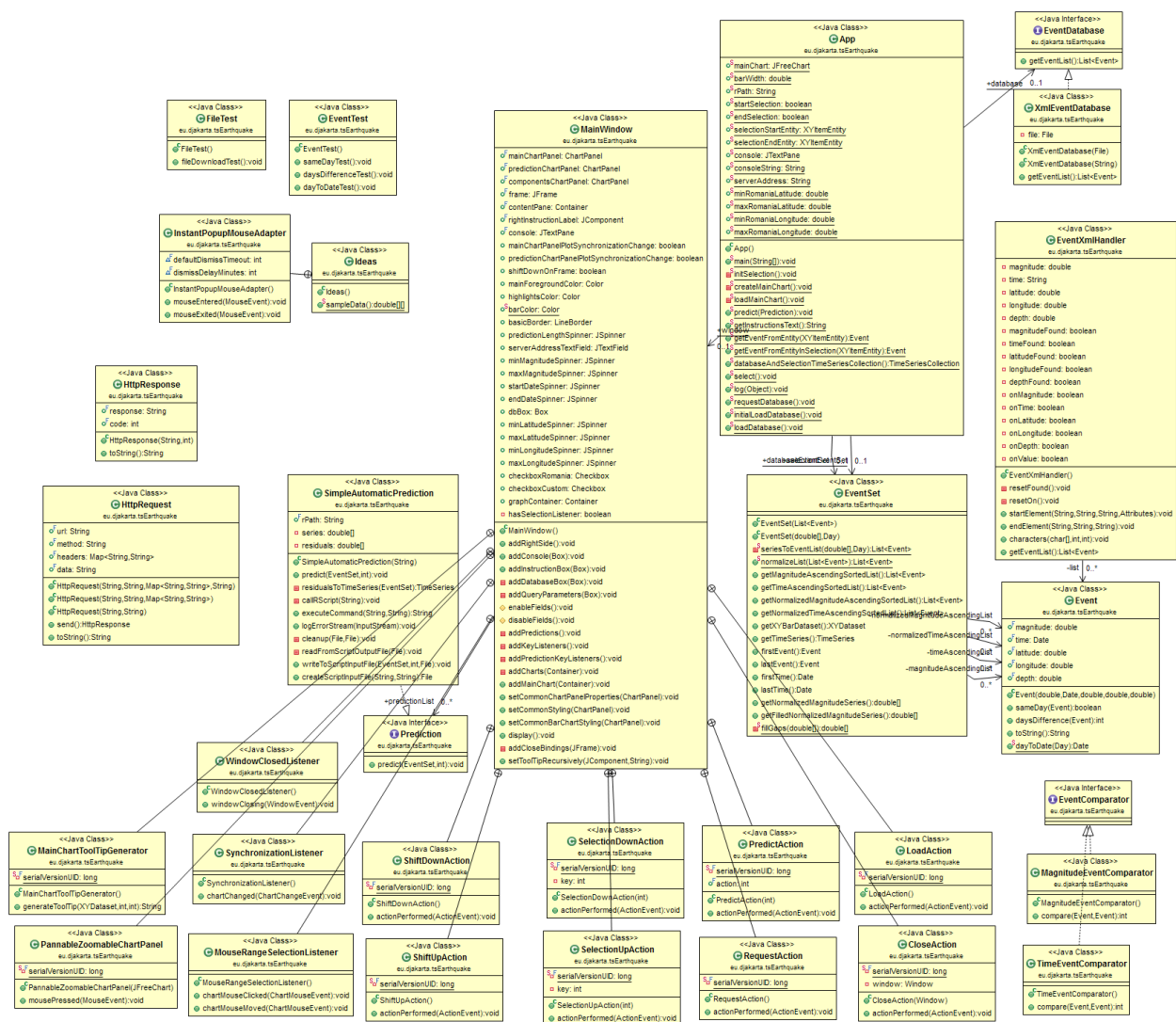
Încărcarea de evenimente din baza de date

Baza de date poate fi împrospătată sau se poate schimba complet prin intermediul zonei de control al bazei de date. Pentru baze de date online de QuakeML, utilizatorul completează câmpurile necesare cererii HTTP și trimite cererea. Se așteaptă descărcarea evenimentelor în baza de date locală și mesajul de succes al descărcării. La terminarea cu succes, utilizatorul poate încărca evenimentele din baza de date locală în program apăsând butonul de încărcare și poate apoi genera predicții noi pe baza noilor evenimente.

Schema internă a aplicației

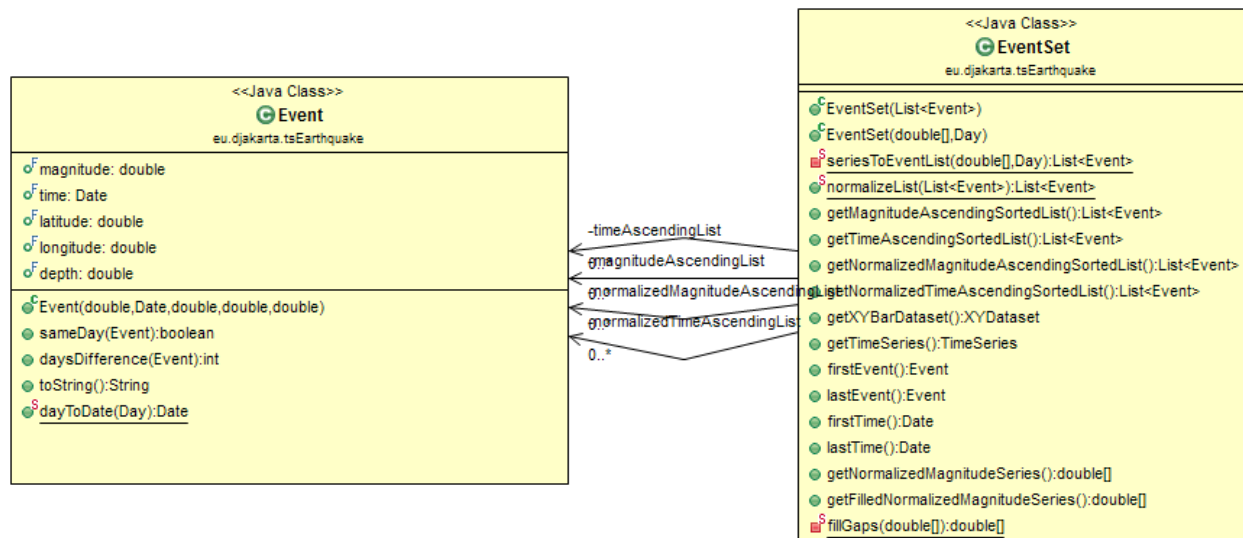
Schema generală

Diagrama completă de clase a aplicației este următoarea:

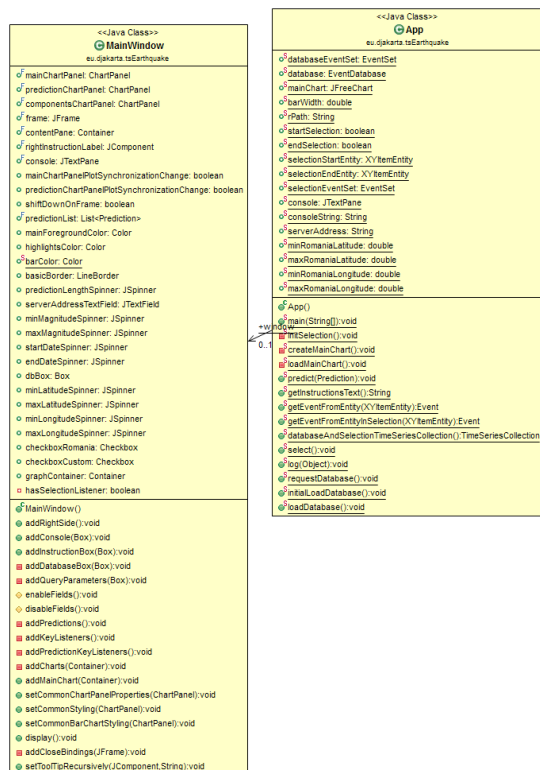


După cum se poate observa din diagramă, cele patru clase principale ale aplicației sunt App, MainWindow, EventSet și Event.

Din punct de vedere conceptual și de structuri de date, aplicația lucrează cu obiecte de tip `Event` și `EventSet`.



Clasa `Event` memorează particularitățile unui anumit eveniment seismic (magnitudine, poziție geografică, moment în timp, adâncime) și oferă funcționalități utile în prelucrarea acestora, mai ales din punctul de vedere al timpului (metodele `sameDay()`, `daysDifference()`, `dayToDate()`). Clasa `EventSet` prezintă metode utile în construirea de serii temporale dintr-un set de evenimente seismice, păstrând datele interne în structuri specifice pentru o performanță ridicată. Astfel, pe lângă metodele de conveniență (`firstEvent()`, `firstTime()`, etc.), clasa întoarce și versiuni salvate local ale unor subseturi/serii temporale/liste ale setului de evenimente (`getTimeAscendingSortedList()`, `getNormalizedTimeAscendingSortedList()`, `getTimeSeries()`, etc.).

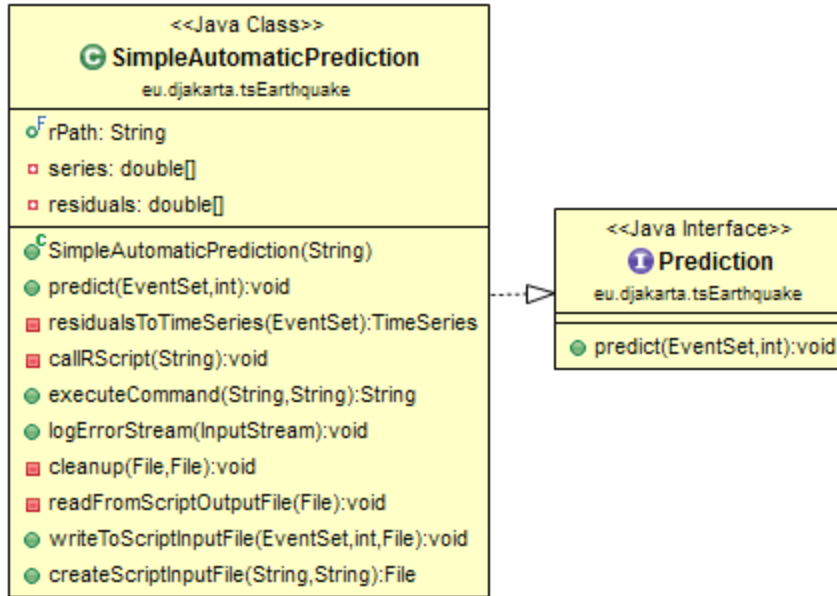


Clasa responsabilă cu logica de bază a modului de lucru este clasa App. Aceasta este și clasa de pornire a programului, având o metoda `main()`. Aceasta reține datele importante aplicației în câmpuri precum `window`, `mainChart`, `selectionStartEntity`, `database`, `databaseEventSet`, `selectionEventSet` și altele. Clasa face legătură între componentele programului de descărcare de baze de date, de citire a bazei de date, de predicție și de creare a ferestrei și a restul interfeței grafice.

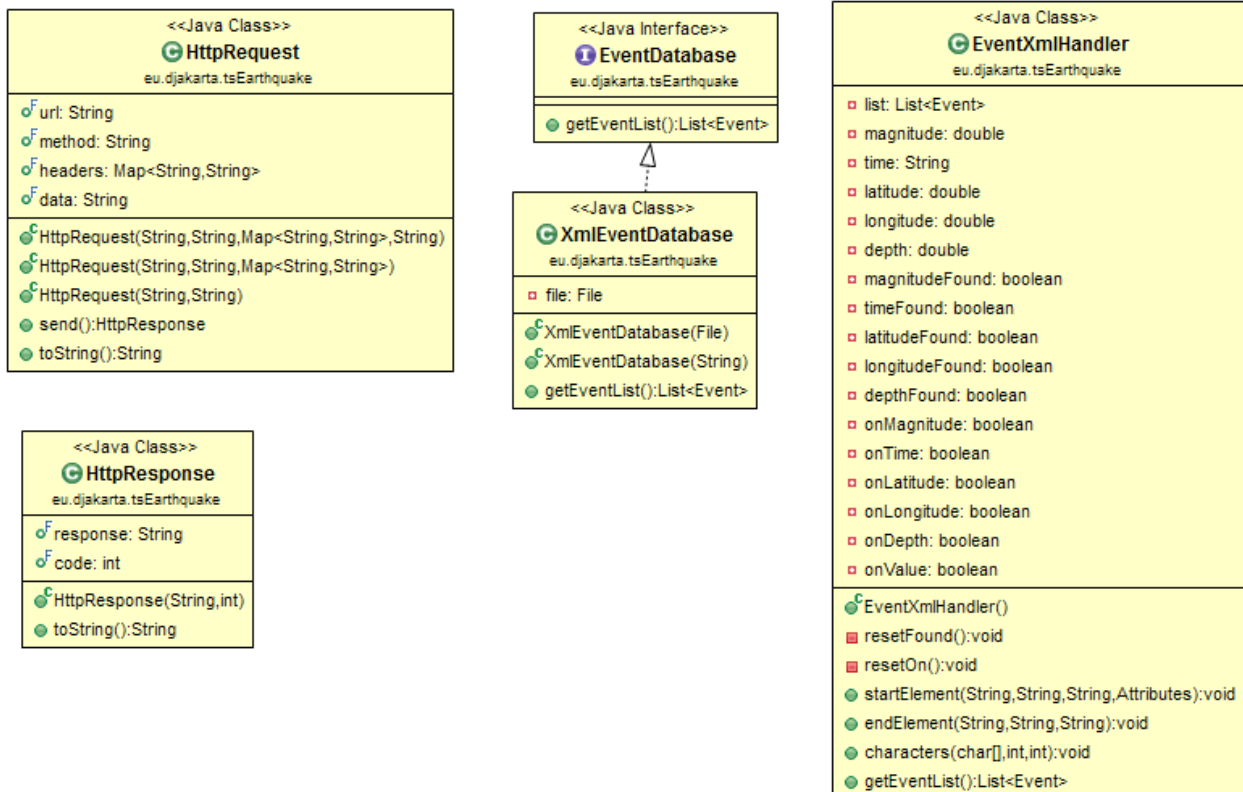
Clasa `MainWindow` are ca responsabilitate afișarea informației și obiectelor de control utilizatorului. Este cea mai complexă clasă, pentru că aplicația necesită un management complex al evenimentelor și elementelor interfeței. Această clasă are mai multe clase interne ajutătoare, în principal pentru tratarea evenimentelor generate de utilizator. Clasa `MainWindow` se folosește în primul rând de tehnologiile Java `Abstract Window Toolkit` și `Java Swing`. Peste acestea, graficele sunt realizate cu ajutorul librăriei `JFreeChart`.

Fluxul de date

Aplicația preia evenimente dintr-o bază de date reprezentată de interfața `EventDatabase`. Aceasta declară metoda `getEventList()`, ce întoarce lista completă de evenimente disponibile în baza de date respectivă. O implementare a acestei interfețe este `XmlDatabase`, folosită în special în cadrul aplicației pentru ușurința în care se pot obține datele despre evenimente seismice în formatul XML.



Predicțiile aplicației se fac prin interfața Prediction. Aceasta declară metoda predictEventSet(), metodă ce primește un set de evenimente, calculează și apoi afișează predicția. Implementarea folosită în cadrul aplicației se numește SimpleAutoPrediction. Aceasta generează un model ARIMA (Autoregressive integrated moving average) potrivit setului de date primit și încearcă să prezică evenimentele seismice din viitor. Intern, clasa SimpleAutoPrediction apelează pentru calculul matematic metode în limbajul de matematică și statistică R.



Cele două interfețe oferă astfel modularitate în fluxul de date al programului. Din perspectiva intrării, datele pot fi primite din orice fel de bază de date de evenimente seismice atâta timp cât programul primește o instanță de `EventDataBase` corespunzătoare bazei de date respective. La fel, din perspectiva prelucrării datelor, aceasta se poate face în orice fel se dorește atâta timp cât programul primește o instanță de `Prediction` cu care să calculeze predicția.

Justificarea soluției și implementării

S-au folosit, în cele mai multe cazuri, practicile de design și cod considerate bune sau standard. Design-ul a fost ales pentru modularitate și extensibilitate. Implementarea interfețelor permite extinderea programului în cele două direcții importante: obținerea de date și prelucrarea acestora.

Pentru interfața grafică s-au folosit Java AWT și Java Swing din considerente de portabilitate și suport. Reprezentările grafice sunt făcute cu ajutorul librăriei `JFreeChart` pentru că oferă un echilibru bun între ușurința de implementare și posibilitatea de setări complexe.

Citirea din fișiere XML este scrisă în standardul Java, fără librării, special pentru fișiere `QuakeML`, pentru o performanță mărită într-o zonă susceptibilă la probleme de viteză. Tot fără librării a fost scrisă și cererea de date de la server pentru un control mai bun al acesteia.

Predicția este o combinație de cod Java și cod R, fapt ce permite o complexitate mare a algoritmilor fără a fi nevoie de implementări specifice.

Ca funcționalități dorite a mai fi implementate amintim afișarea cutremurelor pe hartă, afișarea unei liste de evenimente și filtrarea evenimentelor după anumite criterii, control mai bun în alte aspecte ale prezicerii/selectării/încărcării de evenimente, îmbunătățirea interfețelor oferite de încărcare a datelor și predicție și extinderea modurilor de predicție și analiză în afara limitării seriilor temporale (corelarea cu alte tipuri de date de interes seismologic). Mai punem accent pe dorința de a dezvolta implementări de preluare și predicție mai complexe și mai potente. Nu în ultimul rând, credem că foarte importantă este și rezolvarea problemelor apărute și cunoscute în actuala implementare, momentan mici dar care pot degenera în timp.

Codul aplicației este disponibil sub licență permisivă GNU Affero GPL, liber la contribuții, la adresa <https://github.com/DJakarta/Earthquake-time-series-prediction>.

Concluzii

Aplicația realizată citește date dintr-o bază de date cu evenimente seismice, formează serii temporale și predicții și le afișează grafic. Utilizatorul dispune de controlul parametrilor în majoritatea punctelor de interes.

Problemele întâmpinate și estimarea eronată a timpului necesar pentru a fi rezolvate au făcut imposibilă dezvoltarea de alte facilități planuite. În mare parte, însă, funcționalitatea dorită a aplicației a fost implementată și se speră ca aceasta să se dovedească, dacă nu utilă (din lipsa complexității și funcționalității, în comparație cu alte aplicații similare existente, și, momentan, numărul mic de implementări de interfețe de predicție), măcar demnă de interes.