

METODE DE EXTRAGERE A INFORMAȚIEI DIN CONȚINUTUL WEB ȘI FURNIZAREA ACESTEIA UTILIZATORULUI MOBIL

Emil Stănescu

stanescu@ici.ro

Ileana Stănescu

ileanas@ici.ro

Institutul Național de Cercetare - Dezvoltare în Informatică, ICI, București

Rezumat: Această lucrare descrie unele metode și aspecte practice, referitoare la prelucrarea și folosirea conținutului informațional oferit de Internet pentru utilizatorul mobil. Informațiile oferite acestuia trebuie să fie extrase, de multe ori, dintr-un conținut Web semistructurat, trebuie apoi să fie prelucrate și adaptate pentru necesitățile utilizatorului mobil. Informațiile structurate de tip XML, cu conținut semantic pot să fie o alternativă asupra căreia se fac progrese în cercetările din acest domeniu. O posibilitate de furnizare a informației, la care se face referire în acest articol, o reprezintă folosirea serviciilor Web, care pot fi accesate mai ușor de utilizatorul mobil pentru a-i fi satisfăcute anumite necesități de informare punctuale. În acest sens, se prezintă o aplicație dezvoltată de autori, care furnizează cursul valutar preluat de pe două site-uri diferite, primul dintre site-uri oferind această informație în mod semistructurat (HTML), iar cel de-al doilea site având o formă structurată a datelor (XML). Informația este furnizată utilizatorului mobil prin intermediul unui serviciu Web.

Cuvinte cheie: utilizator mobil, conținut Web, PDA, servicii Web.

1. Introducere

În furnizarea conținutului informațional unui utilizator mobil trebuie să se țină seama de tendințele bazate pe folosirea Web-ului semantic și a serviciilor Web, pe de o parte, și realitatea existentă în momentul de față în Internet unde există miliarde de pagini HTML, care, de multe ori sunt dificil de citit și de interpretat de o mașină care să poată realiza, apoi, prelucrări asupra conținutului lor.

Folosirea facilităților de calcul în orice moment, oriunde, cu orice tip de echipament, a dus la răspândirea termenului de „pervasive computing” care se referă la folosirea unor echipamente de calcul eterogene (telefoane inteligente, asistenți digitali personali – PDA, „tablet PC”-uri, laptopuri și calculatoare staționare de tipul PC-ului). Această gamă variată de echipamente se caracterizează prin configurații și capabilități diferite, precum și prin platforme software diferite.

Multe echipamente mobile sunt mici, atât ca dimensiuni ale ecranului de afișare a conținutului, cât și ca putere de calcul, ca memorie sau ca putere a bateriei.

Din punct de vedere al legăturilor în rețea se pun următoarele probleme:

- *lărgime de bandă*, de cele mai multe ori insuficientă și instabilă pentru o conexiune de calitate;
- *tipuri diferite de canale de comunicație* care încă nu permit folosirea cu ușurință a aceluiași echipament pentru a comuta de la un canal la altul (de exemplu, folosirea unui Smartphone pentru comunicare prin GPRS, WiMax sau Bluetooth);
- *consumul de energie* este diferit în funcție de tipul conexiunii pe care o folosește un echipament mobil (de exemplu, WiMax folosește mai multă energie decât Bluetooth);
- *mobilitatea utilizatorilor* face ca aceștia să folosească diferite echipamente terminale, ceea ce determină o *schimbare a mediului de calcul și a contextului execuției*.

Ca o consecință a acestor limitări și caracteristici, trebuie avute în vedere următoarele cerințe în cazul folosirii pretutindeni a puterii de calcul a unor echipamente:

- **Adaptabilitatea informației la echipamentul clientului.** În funcție de capabilitățile echipamentului se definesc diferite cerințe pentru informația și serviciile care sunt accesibile clientului. În acest caz este necesară transmiterea informației într-o manieră potrivită cu echipamentul mobil respectiv, ținându-se seama că informația care apare într-un browser pe un PC trebuie adaptată pentru a fi accesibilă utilizatorului mobil ce dispune de un echipament cu un ecran mult mai mic, cu o memorie sau putere de procesare mult mai mici.
- **Adaptarea echipamentului mobil la preferințele utilizatorului.** Utilizatorii aceluiași tip de echipament mobil pot avea preferințe diferite asupra modului de folosire a echipamentului.
- **Asigurarea suportului pentru utilizatorul mobil,** prin identificarea lui când el este în mișcare sau în diferite locuri. În acest fel se poate asigura pentru utilizatorul mobil același mediu de informare (de exemplu pagini deja vizitate, documente accesate sau pregătite) și continuitatea serviciilor folosite.
- **Conștientizarea contextului.** Serviciile acordate clientului mobil pot fi de o calitate a informației corespunzătoare dacă se ține seama și de contextul în care se află utilizatorul mobil. În acest sens o

informație importantă este locul geografic pe baza căruia se pot defini caracteristici speciale asupra informațiilor transmise utilizatorului mobil (devierea de la un traseu prestabilit, accesul într-o arie restricționată, găsirea prietenilor care se află în apropiere etc.).

Pentru a adapta conținutul furnizat unui utilizator mobil se poate extinde arhitectura ce folosește 3 nivele (*client - server_de_aplicații - server_de_date*), cu un nou nivel, prin folosirea unui server intermediar, între client și serverul de aplicații, ce furnizează informațiile primare (de bussines). În acest server intermediar se realizează adaptarea conținutului, precum și alte funcționalități cum ar fi localizarea geografică, realizarea de grupuri cu cerințe comune, compresia textului, adaptarea paginilor Web, ștergerea sau comprimarea imaginilor, ștergerea cookies, ștergerea animației GIF, îndepărtarea reclamelor, a applet-urilor și a codului Javascript, a bannerelor, a pop-upurilor, protejarea confidențialității și controlarea accesului. În acest fel utilizatorului mobil i se vor furniza servicii inteligente, adaptate la necesitățile lui.

2. Tipuri de conținut

În articolul „Transactions on knowledge and data engineering” (vezi [7]) sunt considerate trei tipuri de taxonomii pentru instrumentele de extragere a informației în funcție de: *dificultatea unei activități de extragere a informației, tehnicile folosite și gradul de automatizare* (efortul făcut de către utilizator în procesul de pregătire a informației, în corelație cu necesitatea de a folosi sisteme de extragere a informației adecvate fiecărui domeniu).

Conținutul care trebuie prelucrat pentru a se extrage informația poate fi structurat, semistrukturat sau nestructurat.

În prima categorie se încadrează conținutul existent în bazele de date, acesta caracterizându-se prin date care respectă o anumită structură în cadrul bazei de date și în care sunt clar stabilite tipurile lor. Documentele XML pot fi și ele considerate ca fiind structurate deoarece se bazează pe DTD sau o schemă XML.

Datele semistrukturate sunt cele în care se păstrează o structură simplificată cum ar fi log-urile, anunțurile și mesajele în grupuri de discuție etc. Documentele HTML pot fi considerate ca făcând parte din aceeași categorie a datelor semistrukturate deoarece respectă o structură aproximativ regulată.

Conținutul nestructurat este alcătuit din orice texte scrise liber cum ar fi conținutul unor știri, sau din alte tipuri de conținut (imagini, sunet etc.).

Extragerea informației din Web se poate face la nivelul unui site sau a mai multor site-uri, a unei pagini dintr-un site sau la nivel de înregistrare. La nivel de înregistrare se poate face delimitarea acesteia și apoi se pot asocia atribute.

Un exemplu de extragere și folosire a conținutului structurat și semistrukturat pentru realizarea unui serviciu Web a fost realizat de autori în cadrul unui proiect de cercetare având ca beneficiar Ministerul Educației, Cercetării și Tineretului[8]. Acest exemplu este descris în capitolul 5.

3. Fluxul conținutului informațional

3.1. Căutarea subiectelor într-un flux de știri

Printre acțiunile principale care apar într-un proces de căutare a subiectelor în fluxuri de știri enumerăm următoarele: regăsirea documentelor, filtrarea textelor, segmentarea textelor, detectarea noutăților etc.

În proiectul *Topic Detection and Tracking (TDT)* [1] sunt dezvoltati algoritmi de descoperire și de organizare a materialului dintr-un flux de știri, după anumite subiecte. Elementele principale care apar în acest proces sunt:

- *evenimentul TDT*, care este definit ca un lucru particular care se întâmplă la un moment specificat, împreună cu toate condițiile necesare și cu consecințele inevitabile;
- *activitatea* este definită ca un set de evenimente înrudite, care au un scop sau un domeniu comun, care se întâmplă la un moment și într-un loc specificat;
- *subiectul* este definit ca un eveniment sau activitate, împreună cu toate evenimentele și activitățile înrudite direct.

Pe baza acestor elemente se construiește un profil al *subiectului*, format din: *titlul subiectului, evenimentul principal, ce se întâmplă în timpul evenimentului principal, cine este implicat în acest eveniment, când și unde a apărut evenimentul, mărimea subiectului și profilul limbii.*

Acest proces de detectare a subiectelor și de urmărire a lor este semiautomat, fiind necesară și intervenția umană în stabilirea unor subiecte de știri și urmărirea lor.

3.2. Accesul utilizatorului mobil la subiectele semnificative

Pentru accesul utilizatorului mobil la informațiile dorite trebuie să ținem seama de resursele de care dispune acesta și de interesul lui pentru un anumit tip de informații.

În acest sens se poate folosi, așa cum s-a specificat mai sus, un server intermediar care deține anumite informații despre utilizatorul mobil și are posibilități de colectare a conținutului informațional din Web, care să fie, apoi, adaptat pentru a fi folosit de utilizatorul mobil.

Surse de conținut pot fi Bazele de date de întreprindere, fișierele XML, Serviciile Web, Notificările [9], conținutul Web semiestructurat care, eventual, poate conține și informații semantice într-o formă adecvată.

Serverul pentru utilizatorul mobil (SUM) se bazează pe următoarele mecanisme ale infrastructurii [5]:

- *Adaptare conținut* – folosește servicii de adaptare cu ajutorul cărora se furnizează utilizatorilor conținut și servicii personalizate;
- *Furnizare conținut* – folosirea mecanismelor de caching pentru a îmbunătăți furnizarea conținutului bazat pe Web;
- *Generare conținut* – se referă la infrastructura furnizorilor de conținut care găzduiesc versiunile originale ale resurselor și serviciilor.

Fluxul de informații dintre utilizatorul mobil și server poate fi de tip *pull* (utilizatorul solicită niște informații) sau de tip *push* (serverul inițiază operațiunea de transmitere a informației către utilizator pe baza unui abonări anterioare la anumite tipuri de informații sau notificări).

În esență, fluxul de informații dintre un client și server trebuie să cuprindă cel puțin următorii pași:

- a. Clientul emite o cerere sau serverul are o notificare de transmis utilizatorului;
- b. Serverul pentru utilizatorul mobil (SUM) deschide o sesiune utilizator;
- c. În cazul unei cereri din partea utilizatorului mobil i se confirmă acestuia primirea cererii și i se transmite o formă de identificare care trebuie completată pentru autorizarea cererii;
- d. SUM primește conținutul cererii cu datele necesare pentru prelucrarea acesteia.
- e. SUM accesează depozitul cu date despre tipul de echipament și despre caracteristicile personale ale utilizatorului. Acest depozit se poate găsi pe același calculator sub forma unei baze de date relaționale sau sub forma unor date în format XML;
- f. Se fac prelucrările necesare pentru a obține informațiile solicitate și pentru generarea conținutului în funcție de caracteristicile echipamentului terminal și a preferințelor utilizatorului.

4. Folosirea microformatelor pentru reprezentarea informației

Deoarece Web-ul Semantic, care ar fi adecvat regăsirii informației în mod automat, este prea puțin răspândit în milioanele de pagini Web de pe Internet, trebuie luate în considerare și alte abordări. Una dintre aceste abordări are la bază conceptul de *microformat* [3, 4].

Microformatele reprezintă modalități standardizate pentru reprezentarea unor informații care în general se publică. Într-o definiție mai riguroasă, microformatele reprezintă convenții simple pentru încorporarea marcajului semantic pentru un domeniu al unei probleme specifice în documente XHTML/XML, materiale Atom/RSS și în XML, care sunt citibile de om și care normalizează modelele de folosire a conținutului existent prin folosirea unor nume de clase scurte, descriptive, bazate adesea pe standarde interoperabile existente, astfel încât să se permită dezvoltarea descentralizată a resurselor, instrumentelor și a serviciilor [4].

Ca exemplu de astfel de format putem considera *vCard*. Alte microformate se pot găsi în [4].

vCard reprezintă o modalitate standard de realizare a schimbului informației de tip „carte de business”. Acest microformat permite o modalitate standardizată de furnizare a informației care este publicată în mod obișnuit pe Web. Existența standardizată a informațiilor pe Web permite căutările semantice.

Ca exemplu, *vCardul* are următoarele informații:

```
BEGIN:VCARD
```

```
VERSION:3.0
N:Emil;Stanescu
FN:Emil Stanescu
URL:http://sium.ici.ro/
ORG:ICI
END:VCARD
```

Reprezentarea acestor informații în format XHTML este realizată în microformatul *hCard*.

```
<div class="vcard">
  <a class="url fn" href="http://sium.ici.ro/">Emil Stanescu</a>
  <div class="org">ICI</div>
</div>
```

Această reprezentare a informației, în care conținutul informației este separată de reprezentarea ei, permite adăugarea unei foi de stil asociate datelor, care poate permite o reprezentare adecvată în cadrul interfeței utilizator (vezi exemplul de mai jos).

```
<style type="text/css">
  .vcard {border: 1px dotted black}
  .fn {font-weight: bold; text-decoration: none}
  .org {display: inline}
</style>
```

Alte exemple de folosire a microformatelor se pot găsi în [6].

5. Aplicație software: Convertor valutar

Convertorul valutar reprezintă o aplicație de extragere a informației de pe site-uri Web, de prelucrare și de introducere a acesteia într-o bază de date, și realizarea unui serviciu Web care să realizeze conversia pentru schimbul valutar. Din punct de vedere funcțional aplicația poate fi utilizată, de un utilizator, de pe un Pocket PC sau de pe un client de tip PC conectat la Internet. Aplicația are componente care funcționează la nivel server de aplicații (module ale unui server IIS Microsoft) și la nivel server de date (obiecte ale unui server SQL 2000 sau 2005 Microsoft). Aceste componente realizează extragerea informației de pe două site-uri, al BNR și al Băncii CentralEuropene, de unde se preiau cursurile valutilor pentru a permite calcularea schimbului valutar. Conversia se poate face online sau offline. Aplicația a fost dezvoltată utilizând tehnologiile Microsoft Visual Studio, în limbajul C#.

5.1. Arhitectura aplicației

Pentru a răspunde cât mai adecvat cerințelor funcționale ale aplicației, arhitectura aleasă este specifică aplicațiilor dezvoltate pentru utilizatorul mobil, care se încadrează în categoria aplicațiilor multinivel ce au la nivelul de jos sursa de date, conțin unul sau mai multe niveluri intermediare pentru serverele de aplicație, și în nivelul de sus se află clientul. Aplicația include patru module executabile distincte:

- Modulul de preluare de pe site-urile diferitelor bănci a ratelor de schimb și stocarea acestora pe serverul de date;
- Servicii WEB pentru furnizarea de informații privind valutele, băncile, ratele de schimb;
- Modulul client pe pocketPC pentru conversii de valute;
- Modulul client pentru conversii de valute accesibil de pe stații de lucru (PCuri) conectate la rețea.

Interacțiunea modulelor este ilustrată în figura următoare:

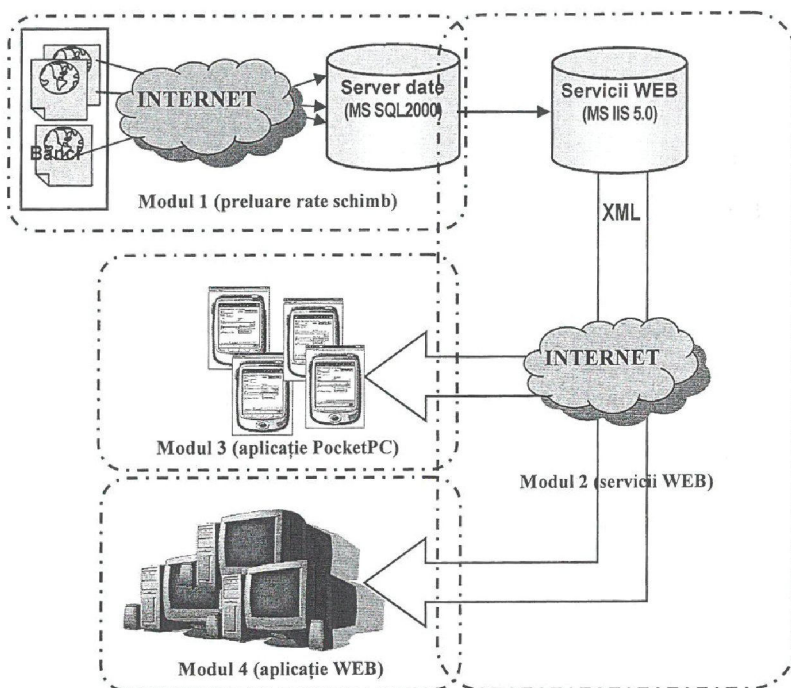


Figura 1. Arhitectura aplicației de curs valutar

Modulul 1 este constituit dintr-un set de funcții de import și extragere a datelor din fișierele făcute publice pe Internet de către bănci. Serverul pentru utilizatorul mobil este folosit pentru stocarea datelor extrase de pe site-urile băncilor ceea ce generează o tratare unitară a informației furnizate utilizatorului mobil, cu toate că sursele de informație sunt deosebite. Lansarea modulului de preluare a informațiilor despre cursul valutar este programată la intervale de timp prestabilite, în fiecare zi.

În prezent, sunt implementate două funcții care transferă ratele de conversie publicate de Banca Națională a României și Banca Central Europeană la adresele www.bnr.ro/Ro/Info/curs_ext.asp, respectiv, www.ecb.int/stats/eurofxref/eurofxref-daily.xml.

Sunt referite și utilizate în C# obiecte din bibliotecile Windows pentru accesarea în Internet și pentru analizarea documentelor HTML, respectiv, XML.

Datele preluate se stochează în baza de date **Valute** pe un server MS SQL 2000/2005.

Legătura dintre aplicația client și serverul de date este realizată prin clasele bibliotecii *System.Data*, namespace *System.Data* și sub-namespace-urile acestuia.

Accesul la pagina BNR - fișier HTML se face printr-un obiect *WebBrowser*. Analizarea conținutului se face utilizând un obiect *HTMLDocument* al bibliotecii *Microsoft HTML Objects Library*, namespace *mhtml*.

Pentru accesarea fișierului XML de pe site-ul Băncii Central Europene este utilizat cititorul XML din .NET, care este un obiect din clasa *XmlTextReader*.

Analizarea documentelor presupune identificarea informațiilor care vor fi stocate în baza de date. În ambele cazuri, documentele sunt parcurse linie cu linie iar în cadrul unei linii sunt identificate coloanele de interes. La sfârșitul operației de analizare a unei linii datele colectate sunt transferate în baza de date.

Pentru completarea bazei de date cu ratele de schimb la zi, aplicația se va executa la momente de timp prestabilite. Programarea lansării automate se face cu utilitarul Windows *Scheduled Tasks*.

5.2. Modulul de servicii WEB

Un serviciu Web XML (pe scurt serviciu Web - *SW*) este o unitate de software programabilă care poate fi accesată prin Internet și utilizată la distanță. Un *SW* poate fi utilizat pe un singur sistem de calcul sau poate fi expus extern pe Internet pentru a fi utilizat pe aplicații multiple. Un *SW* este accesibil printr-o interfață standard, permițând sistemelor eterogene să lucreze împreună ca un singură unitate Web de calcul.

Arhitectura unui proces expus ca un serviciu Web trebuie să aibă un nivel de interfață care să permită consumatorilor să interacționeze cu procesul. Trebuie, de asemenea, să includă un nivel de integrare și transport pentru nivelul interfață. Împreună, aceste nivele constituie arhitectura pentru comunicare. Se disting, astfel, trei componente principale ale procesului: consumatorul (aplicații, obiecte etc), comunicația și serviciul.

Modulul 2 din arhitectura aplicației de conversie valutară, prezentată mai sus, este constituit din serviciul Web *MoneziService*, care este dezvoltat cu mediul *Microsoft Visual Studio (VS)*. Astfel, se poate dezvolta direct *SW* specifice aplicației, infrastructura fiind asigurată de bibliotecile de clase ale spațiilor de nume ASP.NET.

Aplicația asigură gestionarea codului a trei metode, puncte de intrare, ale unui *SW*: *GetCurrencies*, *CurrencyName*, *ConversionRate*.

5.3. Modulul client, de pe Pocket PC, pentru conversii de valute

Aplicația corespunzătoare modulului este dezvoltată în VS, într-un proiect de tip *Smart Device Application*, care utilizează bibliotecii *.NET Compact Framework*.

Aplicația include două grupuri de funcții: 1. pentru conversia unei sume dintr-o valută în altă valută și 2. pentru configurarea preferințelor și actualizarea ratelor schimb stocate pe Pocket PC conform preferințelor.

Aplicația permite utilizarea atât *online* cât și *offline*. Modul online presupune conectarea la Internet și apelul serviciului Web descris în subcapitolul anterior. În acest mod pot fi utilizate toate funcțiile aplicației. Una din ele permite stocarea unei părți a informațiilor obținute *online*. Modul *offline* permite utilizarea informațiilor stocate pentru calcul de schimb valutar.

Informațiile gestionate de aplicație sunt structuri XML. Același format este menținut de la preluarea prin apelul serviciului Web, stocarea unui subset, încărcarea și utilizarea acestuia. Structura XML utilizată este un obiect *DataSet* configurat așa cum este prezentat în continuare.

Interfața utilizator este un formular Windows, specific echipamentului Pocket PC. Formularul posedă două controale de tip tab; câte unul pentru fiecare grup de funcții enumerat mai sus. La încărcarea aplicației se crează două obiecte de tip *DataSet* cu aceeași structură: un tabel (obiect *DataTable*) pentru gestionarea informațiilor referitoare la bănci, un tabel pentru gestionarea informațiilor pentru schimbul valutar specific fiecărei bănci.

Obiectul *ds DataSet* gestionează informațiile necesare conversiei, în funcție de opțiunile utilizatorului. Celălalt obiect *DataSet*, *dsE*, conține informațiile accesate online prin apelul serviciului Web, din care utilizatorul va selecta datele specifice. Acestea sunt copiate în obiectul *ds*. Numai informațiile din obiectul *ds* pot fi utilizate pentru conversia valutară și, eventual, stocate pe pocket PC în fișierul *RateSchimb.xml* din directorul *My Documents\Personal*.

Sunt gestionate informațiile returnate de metodele serviciului Web *MoneziService*, cele mai recente, în raport de data calendaristică specificată la ultimul apel al metodei.

La lansarea aplicației, după crearea obiectului *ds*, se verifică în memoria Pocket PC-ului existența fișierului *RateSchimb.xml*. Dacă fișierul a fost creat ulterior se populează obiectul *ds* cu conținutul acestuia. În figura 3 se prezintă ecranul pocket PC după lansarea aplicației în cazul în care nu există fișierul *RateSchimb.xml* (stânga) și în cazul în care acesta există (dreapta):



Figura 2. Ecrane ce apar la lansarea aplicației de pe Pocket PC

În primul caz, neexistând nici o informație pentru conversie, este activat tabul 2, **Opțiuni**. Dacă există informația stocată, aceasta se încarcă și poate fi utilizată pentru conversii, evident, cu ratele de schimb corespunzătoare opțiunilor de la stocare.

Tot la lansare sunt definite sursele de date pentru controalele de tip liste și comboBox-uri de selecție. Sunt utilizate obiecte de tip *DataView*, care permit filtrarea și sortarea unui subset de date dintr-un *DataTable*. Legătura dintre controlul listă și sursa de date este foarte strânsă, trebuind să existe câte un obiect *DataView* pentru fiecare control listă.

Obiectele *DataView* bazate pe tabele ale *DataSet*-ului *ds* alimentează controalele utilizate efectiv la conversie: **Banci**, **Din**, **În** (din tabul **Conversie**), **Monezi**, **Monezi** **Selectate** (din tabul **Opțiuni**).

Obiectele *DataView* bazate pe tabele ale *DataSet*-ului *dsE* alimentează controalele **Banci** **Existente**, **Monezi** **Existente**.

După lansare, dacă nu a fost stocat fișierul .xml toate listele enumerate mai sus sunt goale. Dacă fișierul există listele bazate pe *ds* vor conține datele corespunzătoare. De asemenea, listele bazate pe *dsE* nu conțin nici un element.

Pentru configurarea preferințelor este necesară conectarea în Internet și apelarea metodei *GetCurrencies* a SW *MoneziService*, se bifează caseta **Conectare**. Parametrul *data* al metodei poate fi specificat completând caseta text **Data**. Implicit, se consideră data curentă. Odată cu popularea *DataSet*-ului *dsE* se va alimenta cu date comboBox-ul **Banci** **Existente** cu lista acronimelor băncilor furnizate de SW, ordonate alfabetic. Lista **Monezi** **Existente** va fi alimentată cu subsetul de date (coduri monezi), ordonat alfabetic, corespunzător primei bănci din listă.

Se pot adăuga/elimina elemente în/din listele din coloana „**Selectate**”. Elementele acestor liste constituie preferințele utilizator și pot fi utilizate la conversie (în tab-ul **Conversie**). Figura 4 exemplifică cele de mai sus:

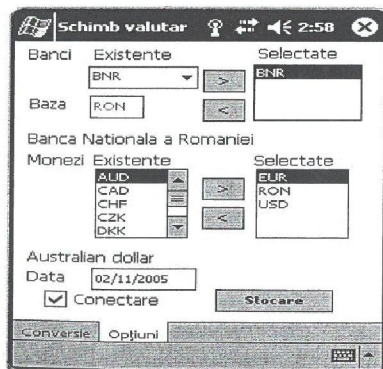


Figura 3. Ecranul de opțiuni

Cu elementele astfel selectate se pot face operații de conversie. Pentru salvarea preferințelor se activează butonul **Stocare**.

Opțiunile se pot utiliza și peste un interval de timp, cu actualizarea ratelor de schimb. Evenimentul declanșat la bifarea casetei **Conectare** actualizează câmpul *rata* al tabelului corespunzător din *ds*.

Operația de conversie valute se poate executa *online* sau *offline*. În primul caz rata de schimb este obținută direct prin apelarea metodei *ConversionRate* a SW *MoneziService*. Lucrând *offline*, ratele de schimb sunt furnizate de fișierul *RateSchimb.xml*.

6. Concluzii

Lucrarea prezintă unele metode de obținere a informațiilor Web și a prelucrării acestora astfel încât să fie accesibile, într-un mod adecvat, utilizatorilor mobili, care au diferite necesități și preferințe de informare și care dispun de echipamente terminale cu caracteristici variate. Se arată că diferitele tipuri de date (structurate, semistructurate sau nestructurate), necesită prelucrări automate sau intervenția umană pentru a se realiza o organizare adecvată a lor. Este prezentată o aplicație care extrage două tipuri de date (structurate și semistructurate) de pe două site-uri Web ceea ce permite realizarea unui serviciu Web de conversie valutară, apelabil de pe un Pocket PC. În continuarea activității noastre vom dezvolta și testa alte aplicații de prelucrare a conținutului informațional din Internet și furnizarea de informații adecvate utilizatorului mobil, folosind diferite canale de comunicație.

Bibliografie

1. * * *: „TDT 2004: Annotation Manual Version 1.2”, August 4, 2004. Disponibil: <http://projects ldc.upenn.edu/TDT5/Annotation/TDT2004V1.2.pdf>.
2. RILLING, P.: Harvesting Web Content into MHTML Archive, <http://www.codeproject.com/useritems/mhtmllib.asp>.
3. * * *: XML and Web Services – Microformats, Last updated Apr 28, 2006. <http://www.informit.com/guides/printerfriendly.asp?g=xml&seqNum=281>.
4. * * *: Microformats Wiki, http://microformats.org/wiki/Main_Page.
5. MAZZONI, F.: Efficient provisioning and adaptation of Web-based services, Ing. Relatore: Prof. Michele Colajanni Anno Accademico 2004 – 2005, DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAZIONE XVIII CICLO Sede Amministrativa Università degli Studi di MODENA e REGGIO EMILIA, (2005).
6. * * *: Exemple de microformate, <http://microformats.org/wiki/hcard-examples>.
7. CHIA-HUI CHANG, M. KAYED, MOHEB RAMZY GIRGIS, K. F. SHAALAN: A Survey of Web Information Extraction Systems, - IEEE Trans. on Knowledge and Data Engineering, Vol. 18, No. 10, October 2006, pp. 1411-1428.
8. STĂNESCU, E., I. STĂNESCU: Utilizarea soluțiilor wireless în asistarea proceselor de afaceri, PN 03.13.05.03, Noiembrie 2005.
9. STĂNESCU, E., V. POPA, I. STĂNESCU: Personalized and Context-Aware Notification and Alerting System for Mobile User, RoEduNet International Conference, Sibiu, 2006.