

# GRADUL DE UTILIZARE AL CARACTERELOR ÎN CADRUL MESAJELOR SMS

Alin Zamfiroiu

zamfiroiu@ici.ro

Institutul Național de Cercetare-Dezvoltare în  
Informatică - ICI București

Academia de Studii Economice București

Carmen Rotună

carmen.rotuna@rotld.ro

Institutul Național de Cercetare-Dezvoltare în  
Informatică - ICI București

**Rezumat:** Articolul abordează serviciul de mesaje text, devenit cel mai accesibil mod de comunicare printre aproximativ 6 miliarde de utilizatori de telefonie la nivel mondial și are ca scop obținerea unor date statistice privind frecvența de utilizare a caracterelor. Multitudinea soluțiilor software precum Nokia PC Suite, Nokia OVI, Microsoft ActiveSync, Apple iTunes, MyPhoneExplorer, AirDroid, face posibilă conexiunea între telefonul mobil și calculator având ca scop interschimbarea de fișiere, urmând ca acestea să fie supuse unei analize pentru determinarea anumitor caracteristici. Soluția propusă în cadrul studiului respectă criteriul interoperabilității prin utilizarea unui sistem portabil de fișiere (CSV). Rezultatele obținute pe baza analizei unui eșantion de mesaje trimise într-o perioadă de 3 ani (perioada de timp prestabilită) evidențiază media caracterelor dintr-un mesaj precum și gradul de utilizare al fiecărui caracter în parte.

**Cuvinte cheie:** mobil, short message service, conexiune, sincronizare, interoperabilitate, export date, caractere, eșantion.

**Abstract:** The article approaches the text messaging service, the most accessible mode of communication among approximately 6 billion phone users worldwide and aims at obtaining statistics on the frequency of characters use. The multitude of software solutions such as Nokia PC Suite, Nokia Ovi, Microsoft ActiveSync, Apple iTunes, MyPhoneExplorer, AirDroid, enable the connection between mobile phone and computer with the purpose of interchanging files, that subsequently will be reviewed in order to determine certain characteristics. The solution proposed in the study complies with the criterion of interoperability by using a portable system of files (CSV). The results obtained by analyzing a sample of messages sent within a period of three years (default period) by a random user, highlights the usability of each character and the average number of characters in a message.

**Keywords:** mobile, Short Message Service, communication, synchronization, interoperability, data export, characters, sample.

## 1. Short Message Service – SMS

Abreviat ca SMS, serviciul de mesaje scurte presupune transmiterea de mesaje de la și către un telefon mobil. Un mesaj scurt nu poate conține mai mult de 160 de caractere. În [1] este prezentată arhitectura serviciului de mesaje scurte, prezentat și în figura 1.

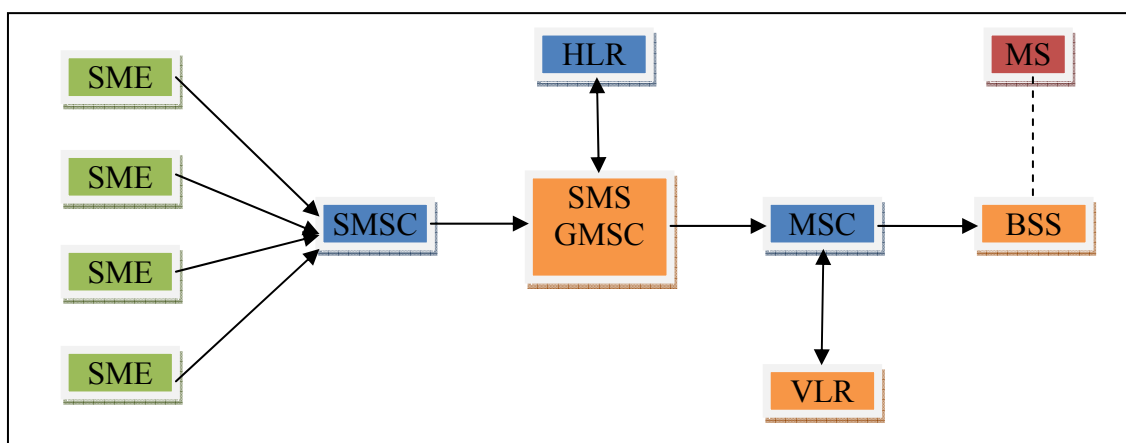


Figura 1. Arhitectura sistemului de mesaje scurte

Primul SMS a fost trimis pe 3 Decembrie 1992 de către Neil Papworth [2]. Inițial SMS-urile nu erau foarte utilizate. În anul 1995 erau trimise 0,4 SMS-uri per utilizator.

Astăzi mesajele de tipul SMS sunt ușor de scris în timpul conferințelor, în timpul prezentărilor sau în timp ce utilizatorii vizionează un film [9]. Astfel, fiecare încearcă să realizeze cât mai multe

lucruri în același timp: să vizioneze un film, să asculte prezentarea din cadrul conferinței și să comunice cu altcineva în același timp prin intermediul mesajelor.

Utilizatorii s-au obișnuit cu tastatura telefonului astfel încât scriu fără a se uita la taste atunci când tastează. Astfel, în timp ce scriu mesaje SMS pot fi atenți la alte activități.

Mesajele SMS sunt de asemenea folosite pentru ca informațiile furnizate să rămână salvate. În cazul în care o persoană este sunată și i se spune o adresă la care trebuie să ajungă, aceasta trebuie să memoreze adresa sau să o noteze în timp ce îi este furnizată prin telefon. În schimb dacă adresa îi este furnizată printr-un mesaj SMS, informația rămâne salvată în telefon până la ștergerea mesajului. În plus, utilizatorul are posibilitatea de a introduce adresa într-o aplicație GPS care prezintă proprietatea de interoperabilitate cu aplicația de mesaje. Aplicația GPS preia adresa din mesaj și furnizează utilizatorului informații despre ruta pe care să o urmeze pentru a ajunge la destinație.

În [3] este prezentată o statistică privind utilizarea mesajelor SMS prin mai multe aspecte precum tipul cuvintelor utilizate și frecvența acestora.

Cel mai utilizat cuvânt în cadrul mesajelor analizate a fost *Du* (traducere din limba norvegiana: *tu*). Al doilea cuvânt cel mai utilizat a fost *jeg* (traducere din limba norvegiană: *eu*). Acestea au fost urmate de prepozițiile *pa* (*la*) și *i* (*în*).

O alta dimensiune avută în vedere este lungimea cuvintelor și mesajelor. Datele analizate arată că femeile trimit în medie mesaje cu 6,95 cuvinte pe mesaj, iar bărbații cu 5,54 cuvinte pe mesaj. Analiza prezentată în [3] arată că doar 6% dintre mesaje conțin abrevieri ale cuvintelor. Aproximativ 82% din mesaje nu conțin litere capitalizate, 11% conțin doar prima literă din mesaj capitalizat, iar 7% au litere capitalizate pentru fiecare frază.

## 2. Nokia Suite

Nokia PC Suite este o soluție software prin intermediul căreia se conectează telefonul mobil Nokia la calculator sau laptop. Folosind acest produs software utilizatorul are acces la informații, mesaje, muzică, contacte, imagini sau videoclipuri stocate pe telefon. Este permisă editarea, sincronizarea și salvarea fișierelor sau a informațiilor existente pe telefon. Conexiunea dintre dispozitivul mobil și calculator se face prin cablu USB sau bluetooth.

Nokia Suite este folosit și pentru instalarea de aplicații pe telefonul mobil, aplicații descărcate din magazinul Nokia OVI, pentru efectuarea de apeluri telefonice sau pentru trimiterea de mesaje.

În figura 2 este prezentată interfața aplicației Nokia Suite.

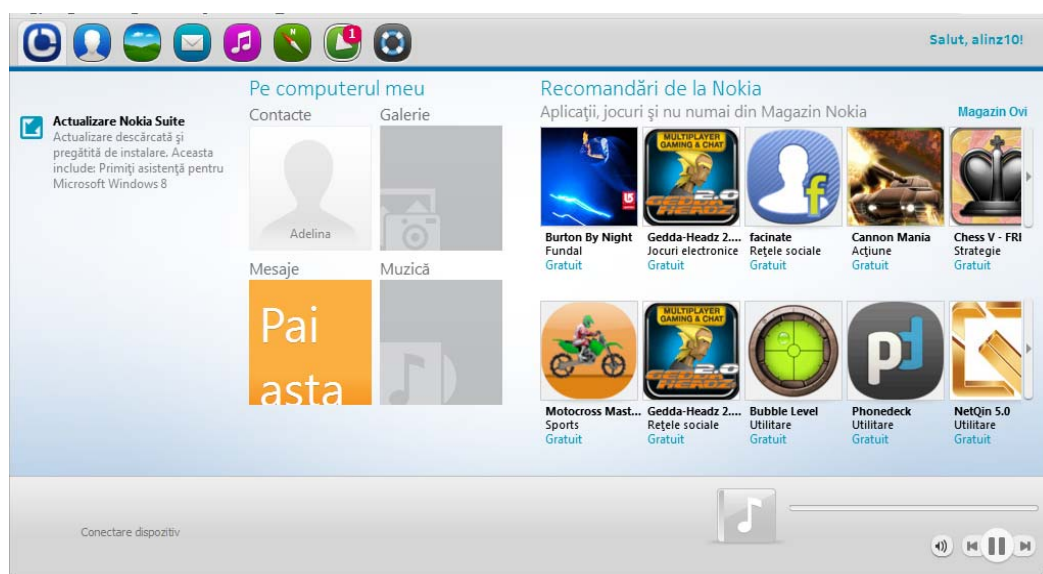


Figura 2. Nokia Suite 3.2.100

În partea de sus a aplicației este prezent meniul care permite accesul rapid la oricare dintre modulele aplicației. În partea dreaptă a ferestrei sunt prezente sugestiile de aplicații din magazinul Nokia pentru a fi instalate de către utilizator pe telefonul mobil.

În partea de jos a ferestrei apare starea conexiunii cu dispozitivul mobil. În figură apare *Conectare dispozitiv*, ceea ce presupune încercarea de conectare la telefonul mobil prin intermediul cablului USB.

Pe dispozitivul mobil, în momentul conectării prin cablu, utilizatorul este întrebat cum dorește să se conecteze la calculator, figura 3. Utilizatorul selectează *Nokia Suite* sau *OVI Suite*, figura 3.

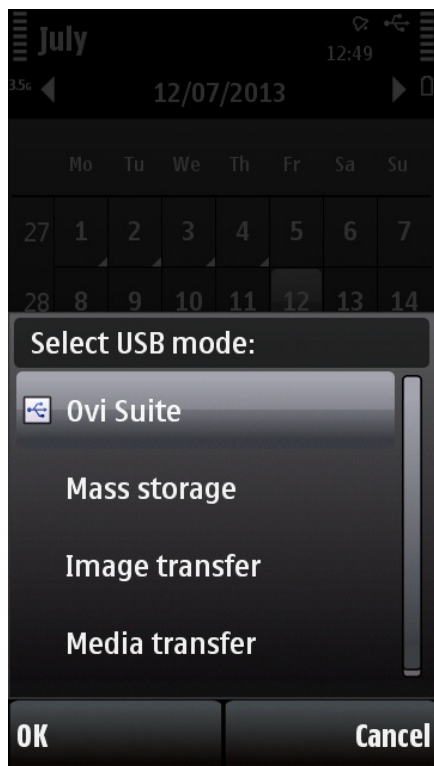


Figura 3. Conectare de pe dispozitivul mobil la OVI Suite

După conectare, în partea de jos a ferestrei aplicației apare dispozitivul care este conectat ca în figura 4, și de asemenea apar și butoanele de sincronizare. În cazul în care aplicația este în modulul de mesaje apare butonul de *Sincronizare mesaje*, figura 4.

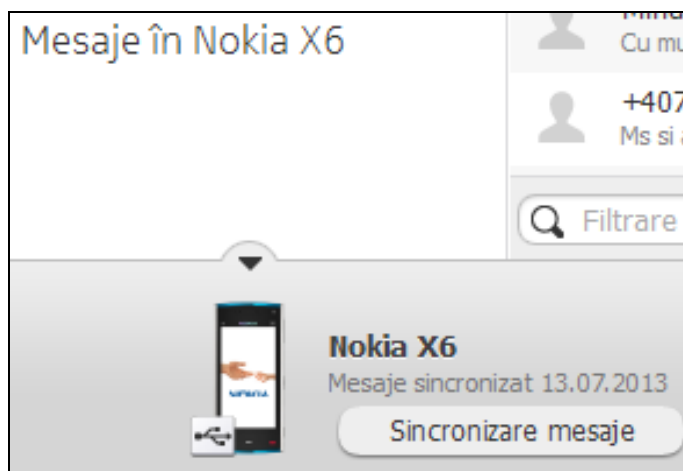


Figura 4. Sincronizare mesaje

Prin apăsarea butonului *Sincronizare mesaje*, sunt salvate pe calculator toate mesajele de pe dispozitivul mobil. După sincronizare, mesajele pot fi accesate fără ca telefonul mobil să fie

conectat la calculator. De asemenea, acestea pot fi exportate într-un fișier extern pentru a fi trimise prin poșta electronică sau salvate pe medii de stocare externe. Pentru exportarea mesajelor se alege opțiunea *Export toate mesajele...* din cadrul meniului *Fișier*, figura 5.

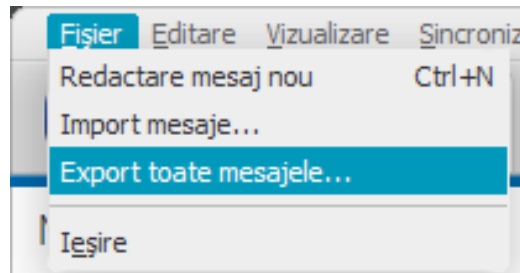


Figura 5. Exportul de mesaje

Prin export, mesajele sunt salvate într-un fișier *csv* și pot fi utilizate ulterior în cadrul altor aplicații sau pot fi importate înapoi în cadrul aplicației Nokia Suite în cazul pierderii acestora.

### 3. Numărătorul de caractere

Se consideră aplicația *LetterCounter* care determină frecvența caracterelor utilizate în cadrul unui text. Pentru determinarea frecvenței sunt memorate într-o variabilă de tip *string* toate caracterele:

```
string litere = "abcdefghijklmnopqrstuvxyz 1234567890:;.,!/?+==-  
_)(*&^%$#@~{ } [ ] / ' ` | \ " " ;
```

Apoi, prin intermediul metodei *countLetters* sunt numărate aparițiile fiecărui caracter:

```
private int[] countLetters(string text)  
{  
    text=text.ToLower();  
  
    for (int i = 0; i < text.Length; i++)  
    {  
        try  
        {  
            counters[litere.IndexOf(text[i])]++;  
        }  
        catch (Exception ert)  
        { }  
    }  
  
    return counters;  
}
```

Textul primit ca parametru este modificat astfel încât să conțină doar caractere mici, deoarece nu se face distincție între caracterele mari și caracterele mici. Apoi se parcurge tot textul, caracter cu caracter și se incrementează în vectorul *counters* frecvența de apariție a caracterului găsit pe poziția curentă în cadrul parcurgerii. La final este returnat vectorul nou creat cu frecvențele de apariție a caracterelor.

Un simplu apel al metodei este realizat pentru prezentarea frecvențelor caracterelor dintr-un text introdus de către utilizator într-un *textbox*:

```
int[] counteri = countLetters(textBox1.Text);
```

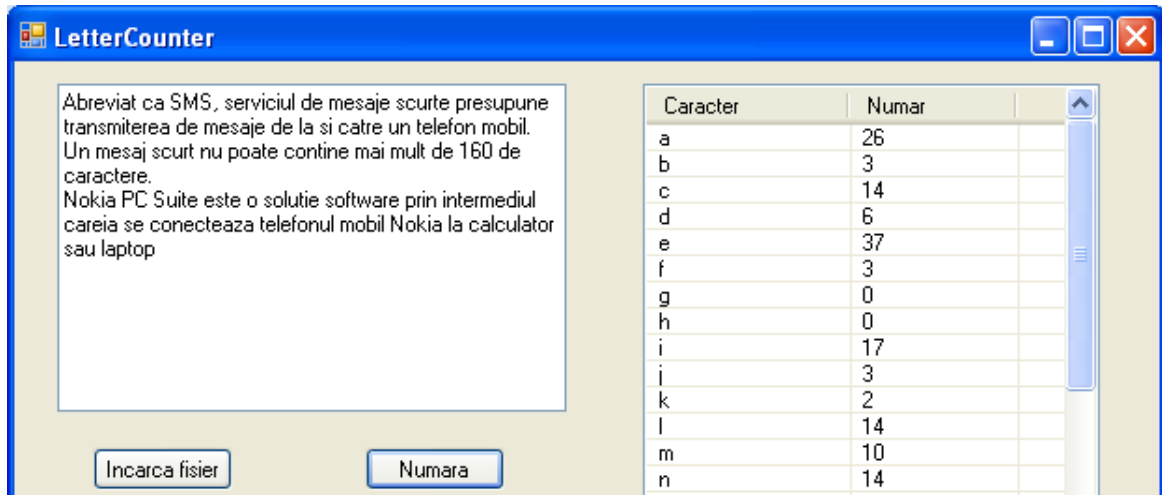
vectorul obținut este prezentat utilizatorului prin intermediul unui control de tip *ListView*, pentru ca acestuia să îi apară sub formă tabelară:

```

for (int i = 0; i < litere.Length; i++)
{
    ListViewItem lv = new ListViewItem(litere[i].ToString());
    lv.SubItems.Add(counteri[i].ToString());
    listViewRezultate.Items.Add(lv);
}

```

Rezultatul este prezentat în figura 6.



**Figura 6. Rezultat tabelar în cadrul aplicației LetterCounter**

Fiind o aplicație nou creată s-a ținut cont de caracteristica interoperabilității cu alte aplicații și astfel a fost implementată posibilitatea de import de fișiere *csv* sau fișiere *txt* pentru analiza mai complexă a unui text mai voluminos sau a mai multor texte.

Interoperabilitatea cu alte aplicații este asigurată prin intermediul fișierelor *csv* care pot fi exportate de către alte aplicații și apoi importate în cadrul aplicației LetterCounter.

În cazul în care se importă un fișier *csv* cu mai multe texte, aplicația LetterCounter furnizează și o analiză asupra numărului de texte și a numărului de caractere din cadrul fiecărui text, figura 7.

Numar mesaj	Numar de caractere
1	35
2	62
3	2
4	55
5	5
6	8
7	12
8	5
9	54

Exporta

**Figura 7. Numărul de texte analizate**

De asemenea, pentru interoperabilitate cu alte aplicații, aplicația LetterCounter prezintă posibilitatea de export a rezultatelor obținute într-un fișier *csv*.

```

StreamWriter file = new StreamWriter("letterCounter.csv");
for (int i = 0; i < listViewRezultate.Items.Count; i++)
{

```

```

        file.WriteLine(listViewRezultate.Items[i].SubItems[0].Text + ";" +
listViewRezultate.Items[i].SubItems[1].Text);
    }
    file.Close();

```

Pe prima coloană este salvat caracterul, iar pe cea de a doua coloană frecvența de apariție a caracterului în cadrul textului analizat. Pentru exemple, din figura 6 au fost exportate rezultatele prezentate în tabelul 1.

**Tabel 1. Rezultat tabelar exportat din cadrul aplicației LetterCounter**

a	26
b	3
c	14
d	6
e	37
f	3
g	0
h	0
i	17
j	3
k	2
l	14
m	10
n	14

Astfel este asigurată interoperabilitatea aplicației, atât pentru datele de intrare, cât și pentru datele de ieșire, acestea putând fi utilizate de către alte aplicații în urma exportului.

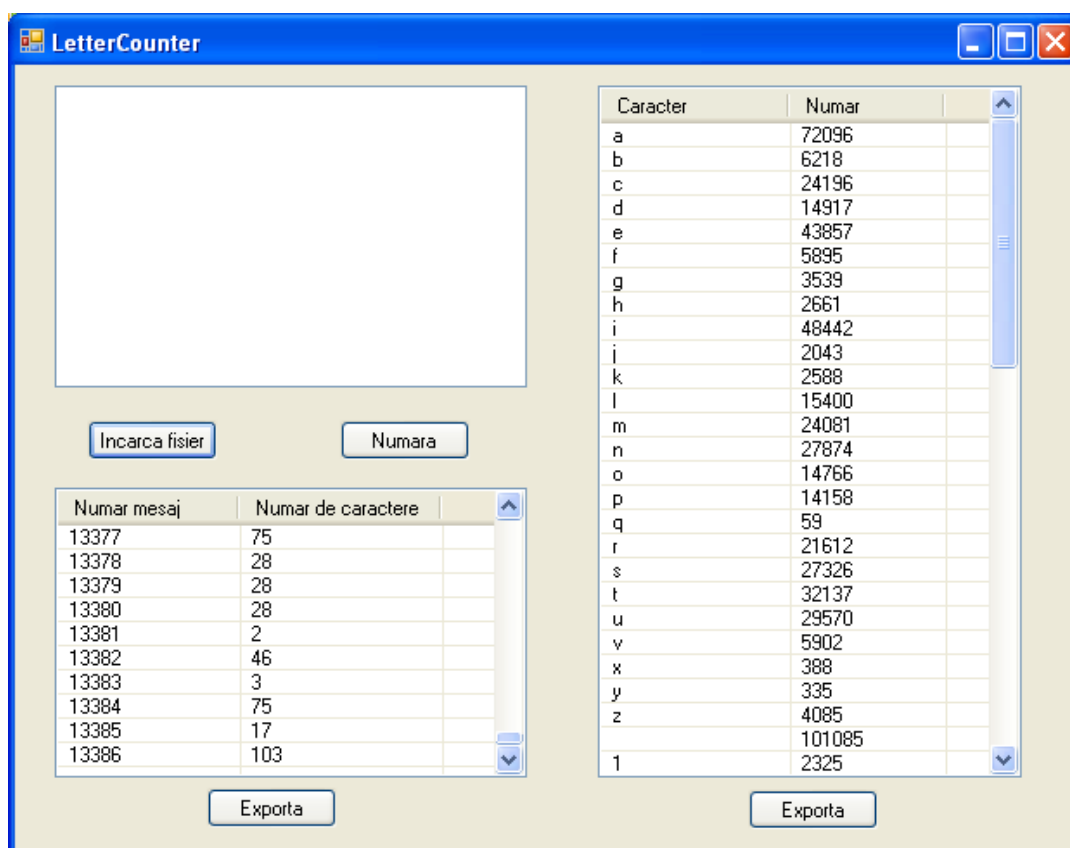
#### 4. Rezultate obținute

Se consideră un eșantion de 13386 mesaje SMS salvate de pe un dispozitiv Nokia X6 prin intermediul aplicației Nokia Suite și apoi exportate într-un fișier *csv*. Mesajele SMS sunt salvate în decursul a trei ani: 23.06.2010 – 12.07.2013 de către o persoană. Sunt exportate atât mesajele trimise cât și mesajele recepționate pe dispozitivul mobil de pe care au fost exportate.

În cadrul fișierului *csv* sunt salvate mai multe informații despre mesajele SMS precum:

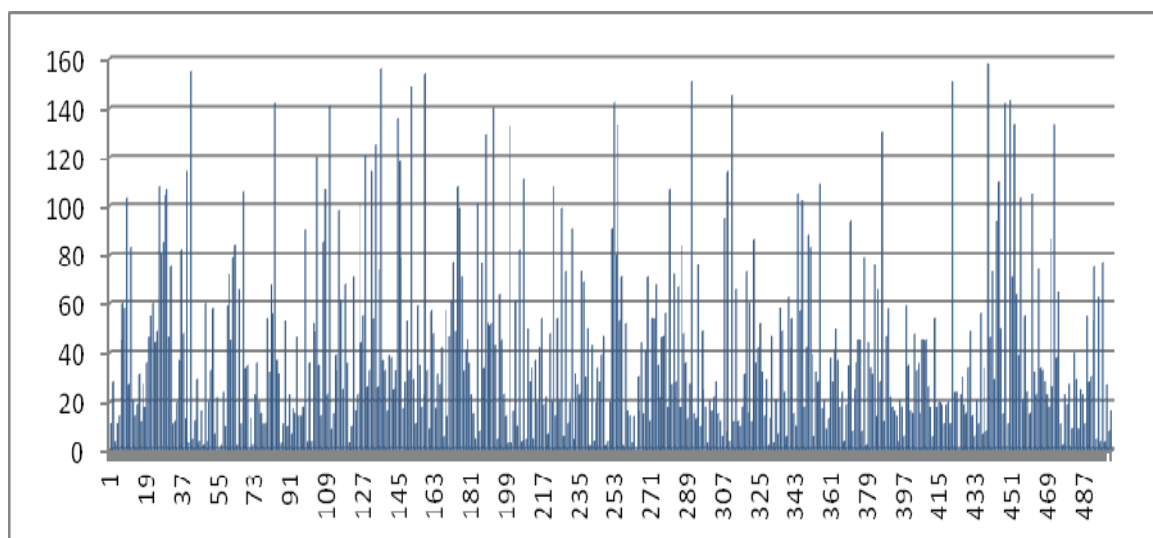
- tipul mesajului: SMS/MMS;
- statusul mesajului: trimis, primit, citit, necitit;
- numărul de telefon cu care a avut loc comunicarea;
- data și ora la care mesajul a fost trimis/primit;

Mesajul text important pentru analiză se află pe cea de-a opta coloană. Astfel, la încărcarea în aplicația LetterCounter trebuie avut grijă ca textul să fie preluat după coloana 8 și apoi trecut prin analiza prezentată în capitolul precedent.



**Figura 8. Rezultate pentru eşantionul de mesaje**

În cele 13386 de mesaje SMS analizate au fost identificate 639575 de caractere cu o medie de 48 de caractere pe mesaj. Pentru 500 de mesaje alese aleatoriu din eşantionul de 13386 este reprezentată grafic în figura 9, lungimea acestora.



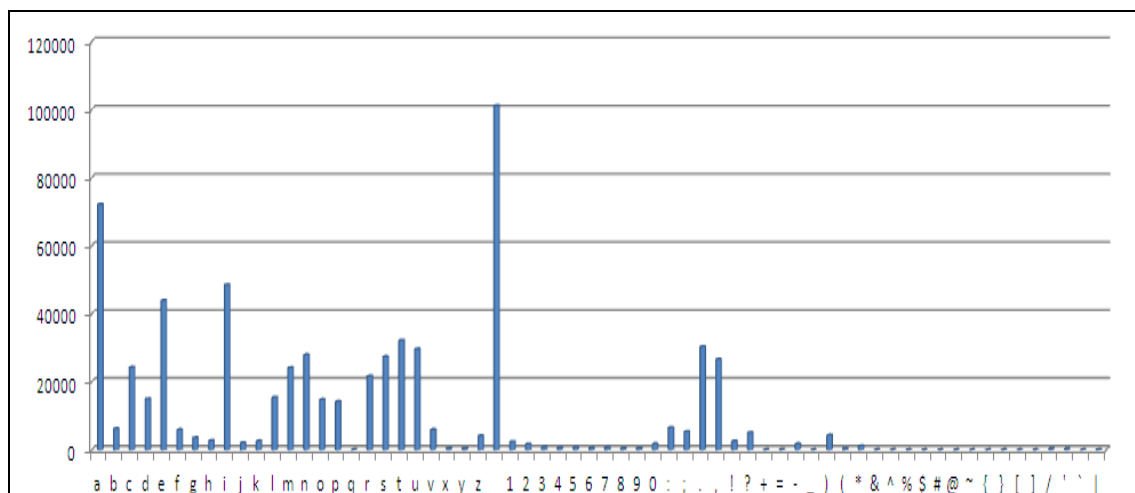
**Figura 9. Lungimea a 500 de mesaje alese aleatoriu**

Cele 639575 de caractere analizate se împart conform tabelului 2.

**Tabel 2. Analiza caracterelor**

Caracter	Frecvența	Caracter	Frecvența	Caracter	Frecvența	Caracter	Frecvența
a	72096	s	27326	:	6459	@	7
b	6218	t	32137	;	5328	~	0
c	24196	u	29570	.	30265	{	0
d	14917	v	5902	,	26598	}	0
e	43857	x	388	!	2513	[	0
f	5895	y	335	?	5021	]	0
g	3539	z	4085	+	8	/	258
h	2661	(spațiu)	101085	=	52	'	249
i	48442	1	2325	-	1714	`	0
j	2043	2	1661	_	6		58
k	2588	3	838	)	4291		
l	15400	4	633	(	319		
m	24081	5	736	*	1104		
n	27874	6	539	&	6		
o	14766	7	700	^	0		
p	14158	8	442	%	6		
q	59	9	474	\$	0		
r	21612	0	1725	#	10		

Aceste frecvențe sunt reprezentate grafic în figura 10.

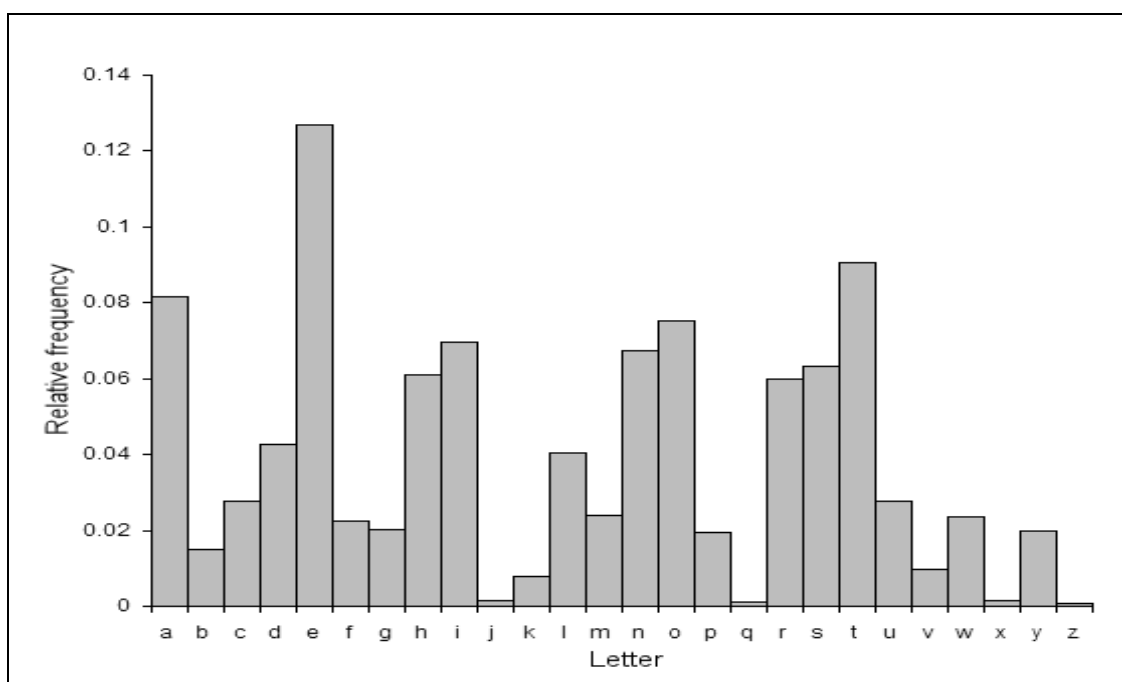


**Figura 10. Frecvența de utilizare a caracterelor în cadrul mesajelor SMS**

Se observă că tasta cea mai utilizată este tasta *spațiu* cu o frecvență de 101085, urmată de tasta vocalei *a* cu o frecvență de 72096.

Frecvența ridicată de apariție a caracterului spațiu este datorată faptului că prin acest caracter sunt despărțite cuvintele. Pe baza frecvenței de apariție a tastei spațiu se determină și numărul de cuvinte din mesajele scurte de tip SMS.





**Figura 11. Frecvența de utilizare a caracterelor [7]**

În [7] este prezentată o frecvență de utilizare a caracterelor în SMS-urile scrise în limba engleză pentru decriptarea acestora. Aici se observă că litera cea mai utilizată este vocala e.

## 5. Concluzii

Mesajele text au devenit una dintre cele mai populare metode de comunicare, acest aspect fiind datorat necesității utilizatorilor de a transmite într-un timp cât mai scurt și folosind minimum de resurse date și informații. Traectoria actuală a acestui serviciu pare să evidențieze faptul că acest tip de comunicare va continua să se extindă prin implementarea unor noi facilități și dezvoltarea celor deja existente. Formatul de date actual pentru mesaje include informații precum: lungimea mesajului, data și ora, numărul de telefon de destinație, formatul, etc.

În cadrul acestui studiu, Nokia PC Suite este folosită pentru exportul datelor de pe telefonul mobil pe calculator într-un format portabil și aplicația LetterCounter implementată în Microsoft Visual Studio 2010 C#, pentru determinarea anumitor caracteristici ale unui text primit ca parametru input. Având în vedere faptul că aceasta este o aplicație nou implementată, s-a ținut cont de caracteristica interoperabilității cu alte aplicații. De asemenea, pentru fișierele importate s-a ales formatul *csv* sau *txt* ce permite operații de analiză mai complexe pentru textele de dimensiuni mari.

Textul primit ca parametru este modificat astfel încât să conțină doar caractere mici, deoarece nu se face distincție între caracterele mari și caracterele mici. Datele de intrare sunt consistente dat fiind faptul că sunt supuse analizei 13386 de mesaje SMS în care au fost identificate 639575 de caractere, înregistrate pe o perioadă de 3 ani. În urma analizei sunt evidențiate date cantitative privind caracterele folosite cel mai frecvent precum media caracterelor dintr-un mesaj și gradul de utilizare al fiecărui caracter în parte.

Datele cantitative obținute vor fi folosite în direcțiile viitoare de cercetare pentru stabilirea modului de construcție a tastaturilor virtuale, astfel încât tastele să fie dispuse în funcție de frecvența de utilizare a caracterelor respective.

## BIBLIOGRAFIE

1. <http://otelecom.net/SMSC.html>
2. <http://neilpapworth.com/>
3. **LING, R.:** The socio-linguistics of SMS: An analysis of SMS use by a random sample of Norwegians, *Mobile communications*, pp. 335 – 349, 2005.
4. <http://www.businessinsider.com/these-are-the-three-most-popular-keys-on-a-keyboard-2013-7>
5. <http://www.imf.org/external/help/csv.htm>
6. <http://nds1.nokia.com/files/support/nam/phones/guides/6370pcsuiteguide.pdf>
7. <http://www.wisegeek.com>
8. <http://www.anuragbhandari.com/2008/12/use-sms-english-to-make-monoalphabetic-cipher-more-secure/>
9. **MENG, X.; ZERFOS, P.; VIDYUT, S.; WONG, S.H.Y.; LU, S.:** Analysis of the Reliability of a Nationwide Short Message Service, *INFOCOM*, pp. 1811-1819. IEEE, (2007)
10. **OLIVER, E.:** Characterizing the Transport Behaviour of the Short Message Service, *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 223-238 ACM New York, NY, USA 2010