

## **Txt2MEDLINE: Text-Messaging Access to MEDLINE/PubMed**

**Paul Fontelo<sup>1</sup>, Fang Liu<sup>1</sup>, Michael Muin<sup>1</sup>, Herman Tolentino<sup>2</sup>, Michael Ackerman<sup>1</sup>**

**<sup>1</sup>Office of High Performance Computing and Communications**

**National Library of Medicine, Bethesda, MD 20894**

**<sup>2</sup>Vaccine Analytic Unit, National Immunization Program, CDC, Atlanta, GA 30333**

### **Abstract**

We developed a text messaging system for processing incoming Short Message Service (SMS) queries, retrieving medical journal citations from MEDLINE/PubMed and sending them back to the user in the text message format. A database of medical terminology abbreviations and acronyms was developed to reduce the size of text in journal citations and abstracts because of the 160-character per message limit of text messages. Queries may be sent as full-length terms or abbreviations. An algorithm transforms the citations into the SMS format. An abbreviated TBL (the bottom-line) summary instead of the full abstract is sent to the mobile device to shorten the resulting text. The system decreases citation size by  $77.5 \pm 7.9\%$ . Txt2MEDLINE provides physicians and healthcare personnel another rapid and convenient method for searching MEDLINE/PubMed through wireless mobile devices. It is accessible from any location worldwide where GSM wireless service is available.

### **Introduction**

Mobile devices, including cellular telephones and personal digital assistant (PDA), are widely utilized by physicians and healthcare professionals [1]. Short Message Service (SMS) or text messaging is available on most digital mobile telephones in use today. It uses the Global System for Mobile Communications (GSM) network, accessible worldwide [2]. Text messaging is one of the most utilized forms of electronic communication. Portio Research estimated that 1000 billion text messages were sent in 2005 [3]. Asia and Europe account for the majority of these numbers. With the predicted doubling of mobile telephones to four billion by 2011 [4], SMS use is expected to increase significantly.

### **Background**

Text messaging is utilized in health and medicine. Reported medical applications include: medication reminders for asthma, diabetes and other patients requiring chronic medication intake, prescription

refills, appointment reminders, smoking cessation programs and dispensing medical advice [5][6]. It is also for informal consultations with medical colleagues.

Physicians need access to reference resources at the point of care to practice evidence-based medicine. Through MEDLINE/PubMed, MEDLINEplus, and other knowledge sources, the National Library of Medicine (NLM) provides portals to current medical literature to anyone with access to the Web, including wireless handheld devices. A text-only Web interface formatted for handheld devices is available at <http://pubmedhh.nlm.nih.gov> for MEDLINE/PubMed. It functions just as well for desktop computers and is suitable for healthcare providers in low bandwidth environments.

The proliferation of mobile phones and the popularity of SMS open new opportunities to extend Evidence-Based Medicine (EBM) resources to mobile phones and other wireless devices. These devices, with even smaller viewing monitors and a 160-character per message limit imposed by the SMS protocol, create challenges to the delivery of appropriate medical content. Although the text-only interface works with mobile phones, one journal citation alone will require multiple text messages. The increased expense and inconvenience reading the messages could become a hindrance to practicing EBM. It can also present difficulties in entering medical terms on a mobile phone keypad.

We addressed these challenges through the development of an SMS system. The components include an inbound/outbound message server with a GSM modem, a medical terms abbreviation and acronym database, abbreviated abstract summaries and an algorithm that reduces the message size.

### **Methods**

#### **Txt2MEDLINE Architecture**

A TER-GX101 TriBand (900/1800/1900 MHz) GSM modem (Round Solutions Ltd) connected to a Linux computer (Red Hat Enterprise) comprises the

Txt2MEDLINE server. A Subscriber Identity Module (SIM) card provides wireless connectivity to the Cingular mobile phone network. UltraSMS (kinks.ultralab.ac.uk/ultrasms) interfaces between the GSM modem and MySQL database.

The inbound/outbound traffic flow is illustrated in Figure 1:

- (1) The mobile device sends a text message to NLM's GSM modem (240-461-7765) through a wireless carrier.
- (2) The SMS center processes the incoming message and forwards it to PubMed.
- (3) Using PubMed's E-utilities, message abbreviation and "the bottom line" (TBL) algorithm, the journal citation is retrieved and processed to the SMS format.
- (4) SMS center sends out the text message through the GSM modem or sends to the user as e-mail.
- (5) Mobile phone receives search result as a text message.

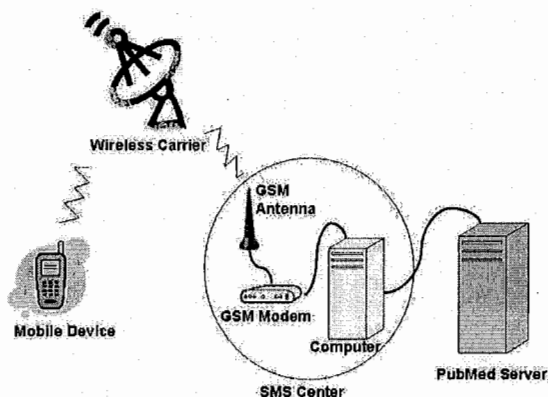


Fig. 1. Txt2MEDLINE Architecture

#### Abbreviation and Acronym Database

Using MySQL, we created an abbreviations and acronyms database. It currently contains approximately 3000 medical terms (see example in Table 1.) We selected the most frequent usage of an abbreviation or acronym in medicine when ambiguity occurred. The database is continuously growing and users are encouraged to submit additions.

Abbreviation/Acronym	Formal Term
CTS	carpal tunnel syndrome
MH	malignant hyperthermia
w/	with
inj	injection

Table 1. Examples of Abbreviation and Acronyms in the Database

#### Incoming Message Processing

Queries are initiated by sending a text message to the Txt2MEDLINE GSM modem. UltraSMS allows the

GSM modem to communicate with the MySQL database.

Search commands are always in upper case text followed by a question mark (Table 2.)

Search Commands	Interpretation
S?	Search
SR?	Send result to mobile device
M?	Send result by e-mail
LR?	Limit result (default=1)

Table 2. Common server commands and meaning

A sample search query format is shown in the box below:

```
S?cts surg vs steroid inj rct M?
username@userhost.com LR?3
```

The server interprets this query as:

```
Search for 'carpal tunnel syndrome'; compare
'surgery' versus 'steroid injection'; retrieve only
'randomized controlled trials' publication types; send
results by e-mail to 'username@userhost.com'; limit
results to 3 articles.
```

Queries may be submitted through messaging services that can send text messages on a computer to mobile phones, such as, AOL Instant Messenger and Yahoo Messenger. The search may be done also by sending an e-mail message through a message service that can convert an e-mail message to a text message to a mobile phone. The Txt2MEDLINE GSM modem can receive data directly from mobile phones.

#### Query Processing

Queries can be sent as a combination of abbreviations and formal terms. Each input term in the query is searched in the abbreviation database. If this input term is found to be an abbreviation, its corresponding formal term will be used in the query. Only formal terms are submitted to E-utilities unless it is an abbreviation known as a MeSH term.

The translated query is sent to PubMed through E-utilities, which are NCBI (National Center for Biotechnology Information) resources for accessing PubMed.

#### Outgoing Message Processing Algorithm

To decrease the size of the original data from E-utilities, we developed a text-abbreviating algorithm that significantly reduces the size of outgoing messages. Its two components are TBL algorithm and the word transformation algorithm.

### **The TBL Algorithm:**

If the abstract is structured or if it contains the word 'conclusion', the segment, sentence or phrase that follows it, will be returned as the TBL. Approximately 9% of citations with abstracts published in PubMed after 1995 are structured or contain the word 'conclusion' in their abstract. All MEDLINE records are included, but citations without abstracts are disregarded.

If no 'conclusion' is found in the abstract, the TBL algorithm parses the journal abstract into sentences identified by punctuation marks (period or question mark.) The process then proceeds as follows: (1) all terms from a 'stop words' list (from PubMed's list) are deleted. (2) A frequency count of the remaining words is made. The top five most frequent words are considered as the key words of the abstract. (3) The sentences in the abstract are then ranked by the frequency of the occurrence of key words. The sentence with the most number of key words present and the last two sentences of the abstract (if they are not the sentences with the most number of key words) will be selected as 'the bottom line' summary. The rationale for choosing the last two sentences is that it often contains significant relevant information useful in summarizing the key points of the abstract. We are currently undergoing a formal evaluation on the validity of these assumptions, although we have found this to be accurate by informal evaluation.

### **Word transformation Algorithm:**

The use of abbreviations and acronyms in current medical literature is high [5]. In developing the database, we were guided by previous studies of methods for systematically abbreviating words and names [6-9]. In previous methods, vowels are regarded as redundant, [8][9] so we eliminated all vowels in the first version of the algorithm. However, the more vowels are deleted, the more difficult it is to reconstruct the original information [6]. We therefore designed a word transformation algorithm that reasonably set the threshold and systematically deletes some redundancy in each word. A trial period allowed refinement of the rules. These rules are still in flux and we continue to revise them based on our experience and as feedback is received.

In general, words with 4 letters or less are not truncated. All consonants are retained. If a word with more than 4 letters is not found in abbreviations database, all vowels are deleted, except when the vowel:

- (1) is the first letter of the word.
- (2) is the first vowel after the first letter of the word.

(3) if vowels occur in tandem, the first vowel is retained.

### **Other Optimization Procedures:**

- Only the last name of the author is retained. When there are multiple authors, a '+' sign is appended immediately after the first author's last name.
- Some of MEDLINE's abbreviated journal titles are further shortened. Abbreviated forms of highly-accessed clinical journals are stored in a database. For all other journals, spaces between words in the title are deleted.
- Publication dates are truncated by using the first two letters of the month and last two digits of the year.

### **Registration Service**

An optional registration service is available to the user. An active authentication step requires sending a verification message from the same phone used to register the account. The user ID is sent thereafter.

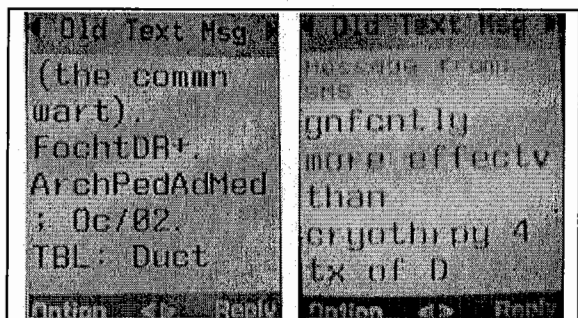
This feature allows the user to send a shorter query. Upon registration, a 5-character account ID will be generated randomly. This account will be associated with a user's mobile phone and e-mail address. Users may then use this account ID to send queries instead of the regular 10-digit phone number and/or e-mail address. This will be especially convenient when users want the results sent to both a mobile phone and by e-mail. For example, instead of sending a message as 'S?mh rct M?username@userhost.com', a registered user can simply send the message such as, 'S?mh rct M?123ab', where '123ab' is the user's account ID. The results will be sent to the mobile phone and e-mail address.

### **Message Truncation Evaluation**

We selected key words from the top 10 most searched terms from *askMEDLINE* [10] to evaluate the system's citation truncation function. Ninety-five citations were tested.

### **Results**

Figure 2 shows an example of text messages received from Txt2MEDLINE. Two text messages were sent for this citation. The user would need to scroll up and down to read the entire message. The query in this example was: 'S?warts duct tape rct'. The result shows one randomized controlled study comparing duct tape occlusion to cryotherapy in the treatment of the common wart.



**Fig. 2. Mobile phone screen shots of a search result.**

Figure 3 and 4 illustrate the results of the message-abbreviating algorithm. The original citation (Fig.3) contains 1783 characters. Since the SMS protocol forces a 160-character limit per message, 12 text messages would be required to send the complete citation. The message-abbreviating algorithm and the TBL algorithm reduced the original citation to 216 characters (Fig. 4) and will only need two text messages.

A randomized controlled trial of surgery vs steroid injection for carpal tunnel syndrome. Hui AC, et al. Neurology, 2005 Jun. BACKGROUND: Decompressive surgery and steroid injection are widely used forms of treatment for carpal tunnel syndrome (CTS) but there is no consensus on their effectiveness in comparison to each other. The authors evaluated the efficacy of surgery vs steroid injection in relieving symptoms in patients with CTS. METHODS: The authors conducted a randomized, single blind, controlled trial. Fifty patients with electrophysiologically confirmed idiopathic CTS were randomized and assigned to open carpal tunnel release (25 patients) or to a single injection of steroid (25 patients). Patients were followed up at 6 and 20 weeks. The primary outcome was symptom relief in terms of the Global Symptom Score (GSS), which rates symptoms on a scale of 0 (no symptoms) to 50 (most severe). Nerve conduction studies and grip strength measurements were used as secondary outcome assessments. RESULTS: At 20 weeks after randomization, patients who underwent surgery had greater symptomatic improvement than those who were injected. The mean improvement in GSS after 20 weeks was 24.2 (SD 11.0) in the surgery group vs 8.7 (SD 13.0) in the injection group ( $p < 0.001$ ); surgical decompression also resulted in greater improvement in median nerve distal motor latencies and sensory nerve conduction velocity. Mean grip strength in the surgical group was reduced by 1.7 kg (SD 5.1) compared with a gain of 2.4 kg (SD 5.5) in the injection group. CONCLUSION: Compared with steroid injection, open carpal tunnel release resulted in better symptomatic and neurophysiologic outcome but not grip strength in patients with idiopathic carpal tunnel syndrome over a 20-week period.

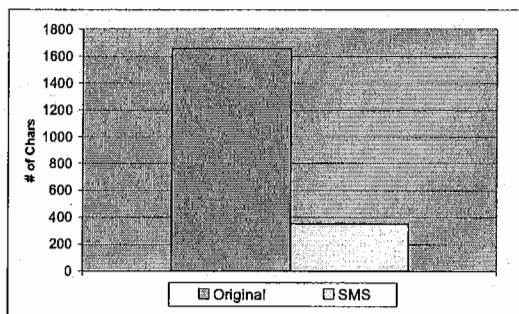
**Fig. 3. Message before Abbreviating Algorithm**

A RCT of Surg vs steroid inj 4 CTS. Hui AC. Neuro. Jun 05. TBL: Comprd w/ steroid inj. open carpal tunnl releas resld N bettr symptomatic & nerphysiologic outcm but not grip strength N pts w/ idiopathic CTS over a 20-week period.

**Fig. 4. Message after Abbreviating Algorithm**

A review of 95 citations to evaluate the message truncation algorithms showed an average of  $1658 \pm 491$  characters per citation in the original citations. The average size after message truncation

was  $352 \pm 114$  characters (Fig. 5.) Therefore, only 2.2 or about 3 messages would be needed, down from the 10.4 messages per citation. Message abbreviating algorithms decreased average message size by  $77.5 \pm 7.9\%$  characters.



**Fig. 5. Message Size Comparison: Original Citation vs. SMS**

We also explored alternate methods for submitting queries to Txt2MEDLINE, other than text messaging by mobile phone. We found that sending e-mail to accounts that will convert e-mail to SMS, AOL Instant Messenger and Yahoo Messenger were successful in submitting queries to Txt2MEDLINE. Multiple tests showed that the average turnaround time was two minutes or less.

### Discussion

Evidence-based medicine and translational medicine depend on convenient access to knowledge sources at the point of care. These resources must be easily accessible and handy. Bottom-line statements are recommended [11].

Two existing conditions favor the alternative access to reference sources discussed in this paper: 1) increased utilization of SMS or text messaging by doctors and healthcare personnel, and 2) doctors' inclination to using abbreviations and acronyms. The proclivity to abbreviations and acronyms starts early in medicine through the use of mnemonic aids in anatomy. The hurried pace during medical school and residency training requires further quick note taking. We have taken advantage of these tendencies with the hope of encouraging the practice of evidence-based medicine through mobile devices.

We initially developed only the abbreviations and acronyms database along with the GSM modem, but it was immediately clear that further truncation of the journal citation was needed to reduce message size. Although cost is a major factor, ease of access and convenience were the prime motivators. Although

there might be a need for it, doctors will not seek references if they are difficult to obtain [12].

Modifications were needed with UltraSMS because of variations in wireless companies' handling of symbols and characters. The European Telecommunications Standards Institute (ETSI) specifies the 7-bit alphabet as the default alphabet for SMS, however variations exist between wireless carriers. The character '@' was the most problematic because it is essential for sending results by e-mail. It required multiple modifications so users can make use of the SMS-to-e-mail method for sending results.

Although real-time messaging services have been successfully tested, they do not work consistently. When a real-time messaging service is used to send a query, it is necessary to specify where the results will be sent, because the system is unable to decode message headers from these services. These headers also vary from service to service and no information is available from the message headers on the identity of the sender.

Several versions of the word abbreviation algorithms are being tested. These modifications will continue based on in-house testing and feedback from users. We are also evaluating the validity of the TBL summary. Feedback from users will guide future modifications of the algorithm

Txt2MEDLINE's architecture could be duplicated in other environments. A local system within a region or country could save on long distance charges, especially for those in overseas locations. Queries can be sent to MEDLINE through the Internet.

## Conclusion

Txt2MEDLINE is an alternative method for searching MEDLINE/PubMed using mobile devices. It is convenient and fast. The service should be available from most areas where a wireless network is available. The SMS center based on a GSM modem, UltraSMS and a Linux computer works effectively for processing queries and retrieving results from MEDLINE/PubMed. The registration process allows the user to send shorter queries with a common ID for mobile devices and e-mail. The abbreviation and acronym database of common medical terms, TBL algorithm and message-abbreviating algorithm transform full-length journal citations into short text messages, suitable for wireless mobile devices. Early user feedback is positive. Refinements of the various algorithms and

evaluation of the validity of the TBL summary are continuing.

## References:

- [1] Fontelo P, Ackerman M, Kim G, Locatis C. "The PDA as a portal to knowledge sources in a wireless setting". *Telemed J E Health*. 2003 Summer; 9(2):141-7.
- [2] Mobile Messaging Future 2005-21010. Global Analysis and forecasts of SMS and MMS Markets. June 27, 2005. Portio Research. <http://www.portioresearch.com>
- [3] Worldwide Mobile Market Forecasts 2-006-2011. Global analysis and forecasts of mobile markets, technology and subscriber growth June 27, 2005. Portio Research. <http://www.portioresearch.com>
- [4] Peersman, C.; Cvetkovic, S.; Griffiths, P.; Spear, H.; "The Global System for Mobile Communications Short Message Service". *IEEE Personal Communications*, Volume 7, Issue 3, June 2000 Page(s):15 – 23.
- [5] Downer SR, Meara JG, Da Costa AC. "Use of SMS text messaging to improve outpatient attendance". *Med J Aust*. 2005 Oct 3;183(7):366-8.
- [6] Ostojic V, Cvorisec B, Ostojic SB, Reznikoff D, Stipic-Markovic A, Tudjman Z. "Improving asthma control through telemedicine: a study of short-message service". *Telemed J E Health*. 2005 Feb; 11(1):28-35.
- [7] Liu H, Lussier YA, Friedman C. "A study of abbreviations in the UMLS." *Proc AMIA Symp*. 2001;:393-7.
- [8] C. E. Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal*, Vol. 27, pp. 379-423, 623-656, July, October, 1948.
- [7] Francis Knowles, "Information Theory and its Implications for Spelling Reform", *Simplified Spelling Society Newsletter*, J2, pp5-13, Spring 1986.
- [8] June A. Barrett, Mandalay Grems, *Abbreviating words systematically*, *Communications of the ACM*, v.3 n.5, p.323-324, May 1960.
- [9] Charles P. Bourne, Donald F. Ford, *A Study of Methods for Systematically Abbreviating English Words and Names*, *Journal of the ACM (JACM)*, v.8 n.4, p.538-552, Oct. 1961.
- [10] <http://askmedline.nlm.nih.gov/ask/otherq.php>
- [11] Smith R. "What clinical information do doctors need?" *BMJ*; 1996 Oct 26 ; 313(7064):1062-1068.
- [12] Finding and applying evidence during clinical rounds: the "evidence cart". *JAMA*. 1998 Oct 21;280(15):1336-8.