Abstract

Computer technologies are inevitably vital in modern and developed society. Without them, we cannot make progress and become competitive in the labor market, not only in Ukraine but also around the globe. Moreover, the communication between employees and even employers is constantly changing. We communicate less formally, use more colloquial phrases. This led to the formation of slang, jargon, and neologisms.

IT slang is extremely popular among programmers, developers, web designers, and other IT specialists in Ukraine. There are not enough studies on Ukrainian IT slang since this industry is rather new for our country and have not been studied yet. In this article, we have considered studies on slang, especially computer slang, provided its definition and developed a Telegram Chatbot which contains a dictionary of IT slang. This study is based on our prior research on lexical and word-formative aspects of IT slang.

Open-source platforms Node.js which executes JavaScript code and MongoDB, non-relational document database, were used to create the chatbot. There are two options of chat implementation. A user can search for a known word or search for a slang term from Noun, Adjectives or Verb categories. The Chatbot also provides the slang word’s definition, its belonging to a part of speech, check the origin of the word, and translation.

Keywords

IT slang terms, Telegram Chatbot, Node.js and MongoDB, Python.

1. Introduction

IT slang is extremely popular with software developers, software QA, web designers, managers, business analysts and other IT professionals. It keeps us with the time, has a social function and marks professionals into a certain group. Information Technology has been growing rapidly, which leads to the development of new terms used by experts in the field. These terms are socially marked and show the evolution in the standard language.

There is a lack of research on Ukrainian IT slang, as this industry to some extent is new for our country and new slangs have been coining constantly. Therefore, the Ukrainian IT market is full of slang terms that are incomprehensible to ordinary people who do not speak English. In addition, there have been no recent sociological attempts to study IT slang. In view of all this, this study will contribute to the study of Ukrainian IT slang from the lexical, word-formation and sociolinguistic aspects and may initiate new studies of sociologists in this field.

The aim of the article is to develop a chatbot in Telegram providing a dictionary where a user can choose a word by parts of speech, check its origin, as well as find its interpretation and/or translation.

The object of research is slang terminology in the IT industry. The subject of the research is the development of a Telegram Chatbot which is done based on sociolinguistic analysis of slang terms, as well as word-forming and lexical aspects of slang in the IT industry. The purpose of the work is to collect slang terms and based on their lexical and word-forming aspects develop a Telegram Chatbot.
The following tasks have been assigned:
1. To define the terms "slang" and "IT slang", their features, functions, and reasons for usage.
2. To analyze existing classifications of IT slang, create own classification and analyze the lexical features of the sample.
3. To create a computer program, the result of which would be a Telegram chatbot with a dictionary of IT slang.

Fuzzy matching and SpaCy have been implemented when developing the chatbot. The aim is to provide an effective execution and perform language processing. Fuzzy matching is a technique that identifies two elements of text, strings or entries. The condition is that they are approximately similar but not exactly the same. This technique is used in computer assisted translation as a special case of a record linkage. In order to find correlations between segments (both texts and entries) in a database of an earlier translation.

SpaCy is an open-source software library which can support Natural Language Processing tasks. They include tokenization, part of speech tagging, named entity recognition, text classification, lemmatization and dependency parsing. Natural Language Processing is becoming very popular nowadays as it ensures with easier interaction with machines compared to the communication with people.

2. Linguistic background of IT Slang

There are a few reasons why IT slang is gaining more and more popularity among IT specialists in our viewpoint: 1. constant development of the latest technologies in the world resulted in creation of new words, which quickly spread among Ukrainian computer scientists and are included in their constant usage; 2. facilitation of communication between employees; they don't need to think long to translate words; slang terms make it easy to understand what is talked about; 3. the difficulty of choosing the correct translation equivalent for an English-language term or its absence; 4. the habit of using IT slang or adopting such vocabulary from colleagues; 5. the desire to impress colleagues, especially new employees, with knowledge of such vocabulary.

Computer slang gives specialists in the field a sense of belonging to a closed society and acts as a basic form of communication. The advent of computer technology and its impact on modern society caused the language phenomena. As a result, both IT employees and people outside computer field use computer slang [1]. I. Shchur claims that Ukrainian IT slang is a kind of heterogeneous phenomenon and singles out two groups 1. those which coincide with units of the common language and constitute the lexical and grammatical basis of slang; 2. slang words and phrases that duplicate officially accepted terms in a language and name things that do not have an officially accepted terminological designation [2].

O. Kovalchuk in her article "Computer slang as a form of youth communication" claims that brevity, shortening of terms and originality are the characteristics of IT slang. She also notes that most neologisms arise in this field, as well as various phonetic, grammatical, and graphic innovations [3].

If we analyze the slang vocabulary of the IT industry diachronically, it can be said that it significantly exceeds the rate of replenishment of the standard Ukrainian vocabulary, as well as many words of professional and group slang. In connection with the rapid progress of information technologies, more and more new words are coined in IT slang, and old ones, on the contrary, have fallen out of use [1; 4]. This tendency in IT slang is explained by extralinguistic factors, as there is a constant replacement of old software and hardware by new ones and is accompanied by the creation of new computer terms.

There is not unified classification of IT terms. Different linguists use classifications that are most suitable for their research or propose their own classification. There is a number of IT Slang classifications by J. Coleman [5], O. Kyrchenko [6, p. 54-55], L. Yaroschuk [7, p. 157-158], M. Bogachuk and D. Bihunov [8] and R. Berest [9].

Since it was only possible to partially analyze the selected slangs according to the existed classifications, we created our own division by thematic feature, considering all aspects of the sample. In total, the sample consists of 289 lexical items, which are mainly collected from a sociological survey (prior conducted) among IT workers, as well as from various forums, blogs, where computer professionals share slang terms that they often use or hear from colleagues. Each group in our classification provides the word in English from which slang word is originated from, Ukrainian slang term, its translation. In the classification, we identified five subject areas, which in turn are divided into groups. Let's consider each subject area in detail and illustrate it with examples.
I. Computer:

a. Functioning, usage, development and modernization of system and application software (34 units).

For example: шоустопери (showstoppers) [shoustopery] — functionality that is added to the product not for the real benefit of the client, but to attract attention to the product; оупен сорс (open source) [open sors] — open software; open-source software; апшика (application) [apishka] — application program interface; коміт (commit) [komit] — to save code changes in the repository (storage). Some words fall under several subject areas. For example, проаедейтимити (update) [proapedeytity] — to update the page/program; bring someone up to speed. We attribute this word to three groups: functioning, usage, development and modernization of system and application software; daily or frequent processes, activities during the working day and communication with customers, team, and management; coordination of work issues.

b. Graphic interface elements (10 units).

For example: дропдаун (drop down) [dropdaun] — a drop-down list; дайлого (dialogue) [dailoh] — dialog box; таба (tab) [teb] — tab; хедер (header) [kheder] — header; block at the top of the site page; фьутер (footer) [futer] — footer; block at the bottom of the site page.

c. Performing simple actions on the computer (14 units): ротейтимити (rotate) [roteitnuty] — to turn the image, element; катнuty (cut) [katnuty] — to cut, скопіпастити (copy and paste) [skopipeistnyut] — to copy and paste text; клікабельний (clickable) [kritakebelyeny] — something that can be clicked with a mouse.

d. Software and hardware problems and their solutions (10 units): за (lug) [lah] — a delay in program execution, lack of response to user actions; міст (migration) — an application to the technical support service; хелпдеск (helpdesk) [helshedsk] — a technical support service, technical support; подебажити (look for) [podebazhty] — the process of finding errors in the code.

e. Using websites, e-mail, and the Internet (12 units). Let us see the examples: нотифікація (notification) [notifikashka] — notifications; заджоититись (join) [zadzhoihtyts] — to join the call; лінка/лінчка (link) [linka/linchka] — a link; форварднути (forward) [forvardnuty] — to forward something, for example, an email.

f. Other (50 units).

Here are the examples: конфіга (configuration) [konfiha] — a configuration file containing all program and system settings; продакшн (product for sale) [prodakshn] — a version of a product (for example, a website, application, or program); ліба (library) [liba]; юай/юайка (UI) [yuai/yuika] — a graphical user interface.

In total, the subject area "Computer" has 130 units and is the most numerous among all, which is explained by the fact that most slang terms are created in relation with software development, from the names of software, applications, new functions, graphical interface, etc.

II. IT services:

a. Words that accompany the process of completing a task, project (54 units). Let us have a look at examples: пул (pull) [pulsty] — to submit a request to make changes from your code branch to the main branch; забеканити (backup) [zabekanyty] — to make a backup copy so that in case of unforeseen situations, the entire system can be returned to its previous state; скек (speka) — a document with detailed description of the requirements, conditions, and technical characteristics of how the functionality under development should work. In this group there are also words that can be attributed to the group "Daily or frequent processes, activities during the working day": шоустопери (showstoppers) [check] — to check, завестімейтити (estimate) [zaestimeityty] — to assess, a forecast of the necessary time or budget for the completion of a task, project; поинвестігейтити (investigate) [poinvestiheityty] — to investigate an issue, a situation, a new product, a function. The word фідбек (feedback) [fidbek] we also include in the group "Communication with customers, team and management; coordination of work issues", as this is the key to the successful completion of the task and satisfaction of the client's requirements and expectations.

b. Output product, final service (11 units): реліз ноутс (release notes) [reliz nouts] — a document describing the changes between the new and previous versions of the software product; реліз (release) [reliz] — the release of a finished product that has passed all stages of development and testing; фронтменд (frontend) — the public part of websites with its help a user can directly contact and interact; юзер гайд (user guide) [yuzer haid] — user manual. This group includes 65 units and is second in number among other groups.

III. Work process organization:

a. Daily or frequent processes, activities during workdays (33 units): стендап (stand-up meeting) [stendap] — daily meetings to discuss the workload; групинити (groom) [hrumyty] — a review of the backlog, see if all tasks are prioritized correctly and which have already been completed; засинкатися/синканутися (sink) [zasynkatysya/synkanutysya] — to give a call to discuss the latest news; зарироититись (report) [zariportytys] — to report; свідчикитись (switch) [svitchtyts] — to switch from one task to another.
(call) [kol]. проапдейти (update) [proapdeityt], оёйлик (daily quest) [deilik] are also in the group "Communication with customers, team and management; coordination of work issues".

b. Communication with customers, team, and management; coordination of work issues (20 units): кик офф митеин (kick off-meeting) [kik of mitinh] — the first meeting with the project team and the project customer; мит ап (meet up) [mit ap] — a meeting of specialists in a certain field to exchange experience or to communicate in an informal atmosphere; пинхувати (ping) [pinhuvaty] — to remind someone about something, to let know.

The subject area "Communication with customers, team and management; agreement of work issues" includes 53 units and is the third in number.

IV. Representatives of IT sector and customers:

a. Hierarchical posts (14 units): трейні (trainee) [treini] — an intern; preferably a graduate of a technical specialty or IT courses; джуних (junior) [dzhun] — a position at the start of IT career. This person has little work experience; ПЕм/проджект менеджер (project manager) [piem/prodzhekt menedzher] — a project manager who runs the product or project and communicates directly with the customer; продакт овнер (product owner) [prodakt owner] — a person who is responsible for the vision of the final product and its value for the user; dew (developer) [dev] — a software developer, a programmer.

b. Colleagues (2 units): тим-менбер (team member) [timmenber]; тімка (team) [tim].

c. Customers (2 units): кастюмер (customer) [kastomer] — a client; стейкхолдер (stakeholder) [steikholder] — a persons interested in the development of a product or project.

d. Other (7 units): контакт-поінт (contact point) [kонтакт-поінт] — a person to whom you can contact a certain question; ріквестор (requester) [rikvestor] — the one who submits a request; свічер (newcomer) [swicher] is a person who came to IT from another field.

Representatives of IT sector and customers group consists of 25 units and is the smallest among the other groups.

V. Other slang words (50 units). We included all words that did not fall under the previous categories to this subject area. Mostly, these are slangs that do not directly refer to software development, daily processes, etc., but denote typical actions, concepts, and qualities. Although all these words have commonly used Ukrainian equivalents, preference is usually given to slang vocabulary. For example, ріферити (refer) [riferity] — to suggest, recommend someone; скед'яз (schedule) [shediul] — a schedule, овервокати (overwork) [overvokaty] — to work overtime; сапортити (support) [saportaty] — to support; to релокейтити (relocate) [rilokaytity] — to relocate company employees to another city or country with a change of place of residence and work; драйвер (driver) [dрайвер] is a person who takes the initiative to manage a project, process or task; бенефіти (benefits) [benefity] — tangible or intangible bonuses used to motivate employees; матчтись (match) [matchtys] — a complete correspondence to something.

3. Analysis of existing systems

We propose to consider well-known solutions for finding a synonyms for IT slang. The most popular systems for processing IT slang include such systems as Slovotvir and Wikislovynx. For a better understanding of these developments, a brief description and their detailed feature description as well as a list of advantages and disadvantages of each given.

Slovotvir, a web application is a platform for searching, discussing, and selecting appropriate matches for borrowed words. The main goal of developing the system was to create a platform for searching, discussing, and choosing appropriate counterparts to borrowed words - one of the characteristic features of "living" languages.

This application has a number of advantages: 1) high base of synonymous substitution; 2) existing community and their involvement in system improvement process; 3) constant updating and support of the software solution. However, it has some disadvantages: 1) the software solution cannot be used in social networks; 2) interaction with third-party applications is limited; 3) lack of possibility of using the product by other developers.

The uk.wiktionary.org/wiki system is a Ukrainian section of the Wiktionary project, a multilingual dictionary and you can adit it. Interpretations and translations of Ukrainian words, as well as translations of words and phrases from other languages, have been collected here and are constantly being updated. However, the main problem of this system is the lack of interpretation of slang concepts, which significantly narrows the functionality of the system. Wiktionary online dictionary has a number of
advantages of the: 1) a huge number of users who can fill the dictionary with new words and keep it up to date; 2) in-depth analysis of the semantic properties of each word; 3) possibility to add and edit the results. However, it contains some disadvantages: 1) limited support; 2) lack of a search function and words of common origin; 3) outdated interface; 4) no ARI for third-party integration.

4. Telegram Chatbot development


The purpose of the article is to create a chatbot in the Telegram messaging program. The user can search for various slangs of the IT industry, check the origin of the slang term, as well as find out its meaning or translation.

The name of the software is "IT slang analysis system".

The main functions of the program are:
1. to search for a word, its meaning and translation.
2. to classify IT slang by parts of speech.
3. to search for slang by categories: noun, adjective, verb.

Chatbot has the following features:
1. The chatbot system works as an application; only one user has access to it.
2. The chatbot can work on any gadget or device provided that the mobile application, web version or PC version of Telegram messenger is installed.
3. The input data is a search slang selected or entered by the user. The output data is information on the interpretation and translation of a slang word.

The database saves user’s all previous messages and shows the current stage. The chatbot is created using a cross-platform Node.js and an open-source platforms MongoDB.

Node.js is a cross-platform which allows JavaScript code to be run outside any browser. JavaScript performs works on the client website, and Node.js - on the server, which allows the latter to create applications without the involvement of other programming languages. Node.js has got some essential features, works with internal libraries, runs commands in JavaScript code, as well as is used as a web server [48]. Another great advantage of Node.js (for creating service applications) is a parallel execution of code, which does not block the execution of other processes [18; 19].

Node.js is used to create services that require constant exchange of data with the user: social networks, chatbots, online text editors, collaboration systems and other applications [18; 19].

MongoDB is an open source cross platform, document oriented database program that does not require a rigid, pre-defined schema [17]. This allows MongoDB to run much faster, scale better, and allow applications to increase their performance.

MongoDB supports the JSON (JavaScript Object Notation) file storage format, which effectively uses data with a complex structure [16; 20; 21].

We used MongoDB to design a database that contains information on step-by-step user guide. JSON File Format Structure contains the following elements:
Ukrainian is the host language of the Chatbot as the target audience is Ukrainian users. It generates communication in Ukrainian through text. It can review past interactions and frame new responses in context. To get started, the user chooses a function `start`.

The following is to choose one of the options: viewing words by categories (noun, adjective, verb) or entering the word manually.

If a user chooses an independent search for the word, they should enter the right one immediately. The system works as one user only application.

The system can be used on any device with an installed application/web version of the Telegram manager. The input data is the search word selected or entered by the user.

The output data is information about the definition of a certain word.
After that, the Chatbot indicates a slang in Ukrainian, its English variant, a brief interpretation of the slang and a translation if one is available for this word. This serves as a response to changing needs and rising expectations, reduce customer wait time and provide immediate answer. It offers customers 24/7 support. The interaction is also easier because customers don’t have to fill out forms or waste time searching for answers within the content. The sample includes 289 unites.

Then the chatbot prompts you to enter the next word to search. To go to the category by parts of speech, the user can click the function "View list of categories". Then they can choose the category "Nouns", "Adjectives", "Verbs", or click "Back". This chatbot is decision tree hierarchies presented to the user in the form of buttons. It reminds the automated phone menus, which we interact on a daily basis and require the user to make selections to find the ultimate answer. This process is shown in Fig.5.
Figure 5: Viewing a section ‘List of categories’
By clicking on the section "Nouns", the user is offered a list of nouns that are available in the database (184 units). This category is the most common.

Figure 6: Viewing a section “Nouns’
To go to another category, you can click the "Back" button. Let us choose the category "Adjectives" which is the smallest and has only 8 units. We make the search according to the procedure mentioned in Nouns category.
Now let's move on to the section "Verbs". It includes 93 units. We repeat the search process as in Nouns category described above.
To go back to the start menu, first press Back to return to the category list, then select Back again to go to the self-search menu.

It has to be mentioned that this chat

4. Conclusions

The idea of creating a Telegram chatbot is rooted in lack of dictionaries and other resources that would collect current slang words in IT industry, their meanings and translation.

The chatbot was created using the Node.js platform, which executes JavaScript code and the MongoDB non-relational document database, which simplifies database management.

The user can search for a word which is known or to choose one of the slang terms from the category "Noun", "Adjective" or "Verb". The chatbot also provides the origin of the slang word, its definition and equivalent translation.

A software implementation of an IT slang chatbot can give the start to the development of new search engines for IT slang and the creation of electronic dictionaries of IT slang. The bot will be useful for people who are interested in Information Technology or work in this field and want to learn more slang terms and figure out their meaning. Although the chatbot covers only slang words selected for our sample, it can be supplemented with new words in the future.

IT slang is modern vocabulary, constantly enriched with new words and is at the peak of popularity in modern IT companies. It also makes some changes to the standard literary language, since a considerable number of slang words have already entered permanent usage among Ukrainians. In this way, IT slang opens new horizons for a comprehensive study of sociolects in the field of Information Technologies.

5. References