

## **TOWARD EXTRACTING DRUG ADVERSE EFFECTS INFORMATION FROM RUSSIAN INTERNET USERS COMMENTS**

Tarasov D.S., Izotova E.D.

Kazan (Volga region) Federal University

*Уникальный код статьи: 5370с4fca078b*

### **Motivation**

Discovery of new adverse drug events, drug interactions and evaluation of efficiency in specific cases in the post-approval period is an important task of the health care system. In recent years new opportunities have emerged to harness data sources that have not been used within the traditional framework.

A huge amount of data is available on the web in form of user reviews and comments but remains largely unharvested due to questionable validity and processing difficulties. In recent study [1] it was demonstrated that user posts on health related websites contain relevant extractable information. However, they still concluded that “user comments remain s a significant challenge to natural language processing system”.

Alternative data sources are especially relevant for drug evaluation in Russian health-care system due to availability of numerous medications not used outside Russia and difficulties in traditional monitoring schemes. However, no natural language processing system in Russian language was developed for this purpose.

### **Results**

We evaluated more that 1500 comments about drug effects in Russian taken from various sources over the internet using hand-annotation and structured web data parsing. From that we conclude that posts on Russian health-related websites and forums also contain information highly relevant for post-approval drug monitoring. Also preliminary studies on automated opinion extraction were conducted on these data, confirming that even simple sentiment analysis tasks on medication-related user comments pose significant challenges and can not be adequately performed using bag-of-word semantical representation, now common in text classification tasks.

### **Conclusions and future work**

Web-scale analysis of user reviews may represent promising new method

for monitoring drug efficiency and side effects in the areas where appropriate statistical data are not available and to complement existing evaluation approaches. However novel language processing techniques are needed to accomplish this task on large scale.

### **References**

Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts in Health-Related Social Networks. Proceedings of the 2010 Workshop on Biomedical Natural Language Processing. 2010. pp. 117-25.

### **References**

1. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts in Health-Related Social Networks. Proceedings of the 2010 Workshop on Biomedical Natural Language Processing. 2010. pp. 117-25.