

Identification of Units and Other Terms in Czech Medical Records

K. Zvára, V. Kašpar¹

¹Center of Biomedical Informatics, Institute of Computer Science, Academy of Sciences of the Czech Republic

Supervisor: Doc. Ing. Vojtěch Svátek, Dr.

Summary

Healthcare documentation in the Czech Republic usually has the form of a free text formatted just using spaces, tabs and line breaks. Extracting information from such a documentation is a challenge that if fulfilled would allow to use Czech medical reports by physicians with no knowledge of the Czech language as well as information transfer to a structured form. It is possible to approach this task as a task of finite-state machine, as a task of the linguistic analysis or as a task of statistics. This article summarizes our findings gained using finite-state machines and using commonly used code lists. Excerpts from real medical reports are translated to English in a way that demonstrates the same or similar problems as in the Czech language. Original Czech excerpts are available in the Czech version of this article.

Keywords: natural language processing, healthcare documentation, medical reports, EHR, finite-state machine, regular expression

1. Czech Healthcare Documentation and Medical Reports

Czech healthcare documentation is usually in the form of a free text. The content of healthcare documentation is regulated by a Ministry of Health ordinance. Legal aspects of healthcare documentation are clarified by e.g. JUDr. MUDr. Roman Žďárek, PhD. in [1]. In the year 2001 the law No. 260/2001 Sb. amending law No. 20/1966 Sb, "on public healthcare", added new paragraphs 67a to 67d regulating healthcare documentation and the National Healthcare Information System (NHIS). This law explicitly states that keeping healthcare documentation as well as its other treatments is considered as processing of personal data according to the law No. 101/2000 "on personal data protection".

According to the law on public healthcare, § 67b section 4, healthcare documentation contains:

- personal data of a patient in the extent necessary for personal identification and history,
- information on a disease, on case history and results of examination and treatment and on other important conditions associated with the health state of a patient and with healthcare delivery procedures.

According to the law on public healthcare, § 67b section 4, records in healthcare documentation must be conclusive, true and perceivable; it should be continuously updated; it must be dated, the author must be identified and it must be signed by its author. All corrections in healthcare documentation must be done by adding new dated entries. Such entries must contain author identifications and authors' signatures. The original record must remain readable.

Appendix 1 of the ordinance No. 385/2006, "on healthcare documentation", regulates the minimal content of healthcare documentation containing information on the examination and the treatment provided. This part called medical report must contain at least: a) information on patient's health state including results of laboratory and other examinations, b) information on treatment provided and patient response, c) recommendations to further healthcare delivery.

Czech medical reports usually bear the form of an unstructured text. The fact that physicians are organized in a professional organization and share a common education implies that medical reports created by different physicians are organized similarly. Especially the order of medical report segments is very common.

The text is usually formatted just using spaces, tabs and line breaks. No markup is commonly used. New medical reports are either created from scratch (starting with an empty text file), from a template (including automatically generated content including patient identification, health insurance agency, physician's name and today's date) or by copying and modifying the last created patient's medical report. The cost of time spent creating the medical reports is usually significantly reduced in cases when reports are created using templates and/or modifying older reports. Therefore creating reports from scratch is not very common. Using old medical reports as templates bears the risk of leaving no longer valid information in the new report. Such problems connected with copying and updating an old report content have been observed also in other countries as described in details e.g. in [2].

Another specificity of Czech medical reports is occasionally inserted a content copied from another sources. Such a content usually originates from a machine of a cooperating unit (e.g. biochemical laboratory). Such an inserted content is sometimes formatted in the form of textual-table, where formatting is retained in case of fixed-width fonts. Data of biochemical examinations are mostly stored in rows, where each cell represents a single result. The common content of such a biochemical analysis data row includes "test name", "detected value", lower and upper limits and a graphical representation of a relation among a lower limit, detected value and upper limit.

Czech medical reports contain large number of typing errors, acronyms and abbreviated words, see e.g. [3]. Used abbreviations usually have a different meaning in different contexts. Punctuation and spacing is also often wrong.

In our study we tried to find out the costs of this initial study compared to positively-screened tests and simulate the costs in the current situation when the screening is not paid by the insurances companies. We also simulated the costs in the situation when the screening could contain only the TSH or only the TPOAb serum tests.

2. Methods

During the 11-13 week of pregnancy, pregnant women were tested for TSH and TPOAb. If the laboratory reported positive results, women were recommended to visit the endocrinologist who eventually started the treatment. Current costs in the insurance system for testing of these three results in the standard test cost 518 CZK (TSH 140 CZK; TPOAb 378 CZK). Since there it is not a standardized method for TSH and TPOAb, each laboratory has developed its own test methods and ranges. In our study we used normal levels used in each laboratory. Orientation normal levels for TSH are 0.1-3.7 mU/l and for TPOAb ≤ 34 IU/ml. From the study all the patients positively screened due to the suppressed TSH were removed to exclude healthy women with physiologic TSH suppression in the first trimester of pregnancy. Therefore we analyzed only the patients with elevated TSH (>3.7 mU/l). We calculated the summary cost for the initial screening and according to the number of positive results of tested parameters we specified the cost of one positive diagnosis.

3. Results

The costs for TSH and TPOAb screening was 1 373 218 CZK. 612 (23 %) patients were tested positive in one of these two parameters. It means the cost for one positive result of 2 243 CZK.

We tested if it would be sufficient to test women for TSH or TPOAb only. Positively tested (up of reference range) for TSH were found 269 women (10.14 %) and from this 77 (28.63 %) where TPOAb positive and 192 (71.37 %) were TPOAb negative. The cost for this screening would be 371 140 CZK; 1 380 CZK for one TSH-positive screening. But 2382 (89.85 %) TSH-screening negative women have 143 (6 %) TPOAb positive results. It means that TSH-only screening may lead to miss 77 patients with TPOAb positive results.

Positively tested for TPOAb-screening only were 220 (8.3 %) women and from this 77 (35 %) have TSH positive results. The cost would be 1 002 078 CZK; 4 555 CZK for one TPOAb-positive screening. But 2 431 (91.70 %) TPOAb negative women have from this 192 (7.9 %) TSH positive screening results. It means that TPOAb-only screening may lead to miss 77 patients with TSH positive results.

4. Discussion

From the results we have found out that both TSH and TPOAb screening tests are not substitutable and they have to be screened together although the costs for the TSH screening is 2.7 times most cost-effective than TPOAb screening. Screening of one of the above parameters could mean lost of 77 positive women.

Additionally, screening of these two parameters is important due to fact that positive results of each parameter mean for the mother and child a different risk: Positive TSH can cause improper development of child's brain during the pregnancy and positive TPOAb results can cause the post partum thyroid disease.

Acknowledgements

This work was supported by a grant from the Czech Health Ministry IGA NS10595-3 and by the project SVV-2010-265513.

References

- [1] Hollowell J.G., Staehling N.W., Flanders W.D., Hannon W.H., Gunter E.W., Spencer C.A., Braverman L.E.: Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab.* 2002 Feb;87(2):489-99.
- [2] Springer D., Zima T., Limanova Z.: Reference intervals in evaluation of maternal thyroid function during the first trimester of pregnancy. *Eur J Endocrinol.* 2009 May;160(5):791-7. Epub 2009 Feb 19.
- [3] Leung A.S., Millar L.K., Koonings P.P., Montoro M., Mestman J.H.: Perinatal outcome in hypothyroid pregnancies. *Obstet Gynecol.* 1993 Mar;81(3):349-53.
- [4] Haddow J.E., Palomaki G.E., Allan W.C., Williams J.R., Knight G.J., Gagnon J., O'Heir C.E., Mitchell M.L., Hermos R.J., Waisbren S.E., Faix J.D., Klein R.Z.: Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med.* 1999 Aug 19;341(8):549-55.

- [5] Man E.B., Brown J.F., Serunian S.A.: Maternal hypothyroxinemia: psychoneurological deficits of progeny. *Ann Clin Lab Sci.* 1991 Jul-Aug;21(4):227-39.
- [6] Dosiou C., Sanders G.D., Araki S.S., Crapo L.M.: Screening pregnant women for autoimmune thyroid disease: a cost-effectiveness analysis. *Eur J Endocrinol.* 2008 Jun;158(6):841-51.
- [7] Grosse S.D., Matte T.D., Schwartz J., Jackson R.J.: Economic gains resulting from the reduction in children's exposure to lead in the United States. *Environ Health Perspect.* 2002 Jun;110(6):563-9.
- [8] Klein R.Z., Sargent J.D., Larsen P.R., Waisbren S.E., Haddow J.E., Mitchell M.L.: Relation of severity of maternal hypothyroidism to cognitive development of offspring. *J Med Screen.* 2001;8(1):18-20.
- [9] Sonnenberg F.A., Beck J.R.: Markov models in medical decision making: a practical guide. *Med Decis Making.* 1993 Oct-Dec;13(4):322-38.

Contact

Mgr. Zdeněk Telička

Third Department of Medicine,
General University Hospital and First
Faculty of Medicine, Charles University
Prague,
U Nemocnice 1
128 08 Prague 2
Czech Republic
e-mail: zdenek@telicka.cz