

УДК 613.955/.96-07:004.853:007

DOI: <http://dx.doi.org/10.11603/mie.1996-1960.2016.1.5941>

СТРУКТУРИЗАЦІЯ МЕДИЧНОЇ ІНФОРМАЦІЇ ЗА ДОПОМОГОЮ ОНТОЛОГІЇ

В. В. Приходнюк

Інститут телекомунікацій та глобального інформаційного простору НАН України

Представлено методологію структуризації медичної інформації, що надходить з різних джерел, з подальшим формуванням онтологій. Структуризація проводиться для неструктурованої (природномовні тексти) інформації або наборів слабкоструктурованої інформації (великі набори різнорідних таблиць).

STRUCTURING OF MEDICAL INFORMATION USING ONTOLOGY

V. V. Prykhodnyuk

Institute of Telecommunications and Global Information Space of NAS of Ukraine

This article describes methodology of structuring of medical information, which than can be used to create ontologies. Structuring is possible for unstructured (natural-language texts) or ill-structured (large sets of non-consistent tables) information.

Вступ. Наш час характеризується постійним зростанням обсягів інформації. Відповідно, зростає кількість слабко- і неструктурованої інформації, зокрема, природномовних текстів. Оброблення великих обсягів такої інформації людиною часто буває неефективною, а іноді взагалі неможлива. Виникає необхідність в структуризації такої інформації, що значно спрощує її сприйняття людиною, дає можливість відшукувати протиріччя, невідповідності і неоднозначності, а також має інші переваги.

Найпростішим способом структуризації інформації є її подання у вигляді таблиці, але для складних структур даних (а саме такі, як правило, виникають при обробці великих масивів даних) це часто буває незручно. Тому значно ефективніше подання цих даних у вигляді однієї або кількох пов'язаних між собою онтологій.

Результати й їх обговорення. Основними етапами в структуризації тексту є виділення з нього термінів і визначення взаємозв'язків між цими термінами.

На кожному з цих етапів необхідно оперувати певними знаннями про мову, якою написано текст, що обробляється. Особливо важливе це завдання в таких галузях як медицина, де є велика кількість спеціалізованих термінів. Як правило, термін є іменною групою (приміром, «Лимфатическая система»). Речення, які містять важливу інформацію, що може стати частиною онтології, як правило, складаються з кількох тер-

мінів і ключових слів, що вказують на характер взаємозв'язку між даними термінами. Наприклад, речення «Лимфатическая система состоит из лимфатических капилляров, лимфатических сосудов и лимфатических узлов» задає структуру, показану на рис. 1.



Рис. 1. Фрагмент онтології, побудований на основі одного речення

Крім природномовних текстів, важливим джерелом інформації є слабкоструктурована інформація, наприклад, набори таблиць.

При обробці наборів таблиць основним завданням є ідентифікація полів, які вказують на ієрархію. Наприклад, таким полем може бути індекс: об'єкт з індексом «3.1» буде дочірнім для об'єкта з індексом «3». Приклад перетворення таблиці в онтологію показано на рис. 2.

Всі інші поля задаватимуть властивості об'єктів онтології. Ці властивості формують певний контекст об'єкта, і на їх основі можна виконувати операції, такі, як багатокритеріальне ранжування. Але найважливішою перевагою такого підходу є можливість пов'язування об'єктів, що належать

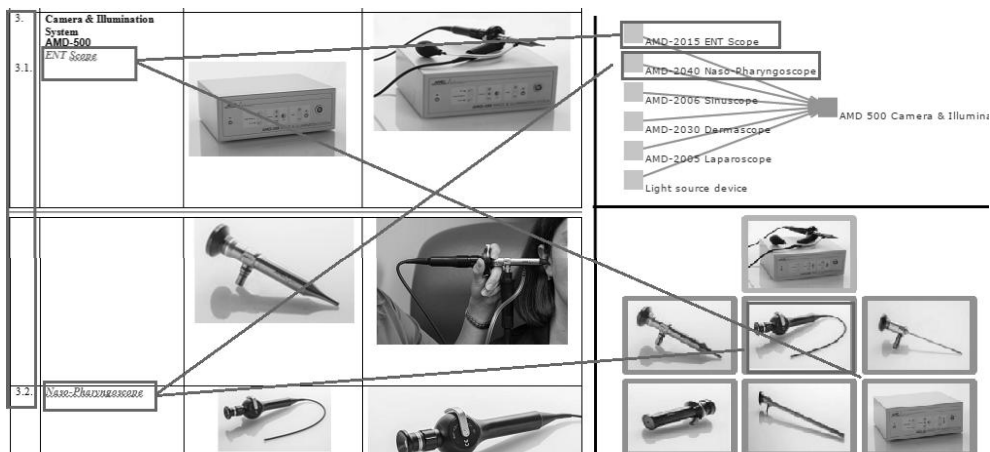


Рис. 2. Приклад перетворення таблиці в онтологію

до різних онтологій, наприклад, за рахунок індексованого пошуку.

Висновки. Описана методологія структуризації інформації може використовуватись для обробки

великих масивів даних. Створене на її основі програмне забезпечення успішно працює при обробці медичної, географічної та інших видів інформації.