ClinicalKey: Terminology driven Semantic Search

Sivaram Arabandi[1*], Helen Moran[1]

[1] Smart Content Strategy, Elsevier Health Sciences, USA

1 INTRODUCTION

In this Information Age, we have a variety of sources to fulfill our information needs. More and more of the data is available online, and search engines such as Google, provide us with powerful tools to find specific pieces of information. However, the data is growing exponentially resulting in an Information Overload [Bergamaschi & Guerra]. This phenomenon is all too common in the healthcare domain too where clinicians are spending increasing amounts of time filtering out useless information to find what they are looking for.

ClinicalKey is an innovative new online resource built on Elsevier's *Smart Content* – searchable journal, book, image and video content tagged to *EMMeT* (*E*lsevier *M*erged *Med*ical *T*axonomy). It is designed from the ground up to provide improved access to clinical information, providing comprehensive, trusted clinical answers quickly (Figure 1).

2 SMART CONTENT PLATFORM

2.1 EMMeT

EMMeT is a clinical terminology model that is being developed to serve as an authoritative reference for clinical terms and the relations between them. *EMMeT* is envisioned as a multi-product, re-usable ontology resource i.e. it will serve the needs of multiple applications. It is based on UMLS, and as of March 2012, has over 1 million concepts and 3 million synonyms. The concepts originate from a subset of UMLS terminologies mainly SNOMED CT, RxNORM, ICD-9, and CPT; from Gold Standard drug database; and a number of terms have been introduced locally for aiding search.

2.2 Smart Content

Smart Content is content with a high level of structure, created by annotating text with a standardized terminology. The terminology, with its logical structure, adds the semantic meaning – what the content is about and how the different pieces of content relate to each other.

Rindflesch and Aronson describe the use of Natural Language Processing (NLP) with UMLS to extract usable semantic information from Medline abstracts. In creating *Smart Content*, natural language processing is used by a *Query Parsing Engine (QPE)* to identify concepts in Elsevier's medical corpus consisting of more than 400 top journals, over 700 books and multimedia, as well as expert commentary, MEDLINE abstracts and select third-party journals. Once the *QPE* has identified the term labels – synonyms, acronyms, abbreviations, etc. in *EMMeT* – mapping rules are applied by a *Concept Mapper (CM)* to create a searchable index of concepts with relevancy scores and links to the originating documents. The resulting concept index is also used to generate RDF satellites that populate Elsevier's *Linked Data Repository (LDR)*.

2.3 ClinicalKey

The first product to utilize Elsevier Smart Content, *ClinicalKey* is a web-based application for clinicians in the hospital setting. It provides a simple search interface for users to enter text that is then processed by the *QPE*. The *QPE* interprets the search text and suggests *EMMeT* terms for auto-complete [Figure 2]. Along with the term labels, additional information such as the Semantic Type of the term is also displayed for disambiguating similar sounding terms. Furthermore, the term-level relations from *EMMeT* are used
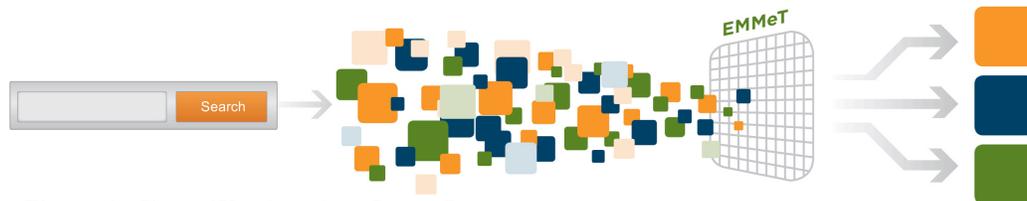
**Figure 1:** ClinicalKey based on Smart Content
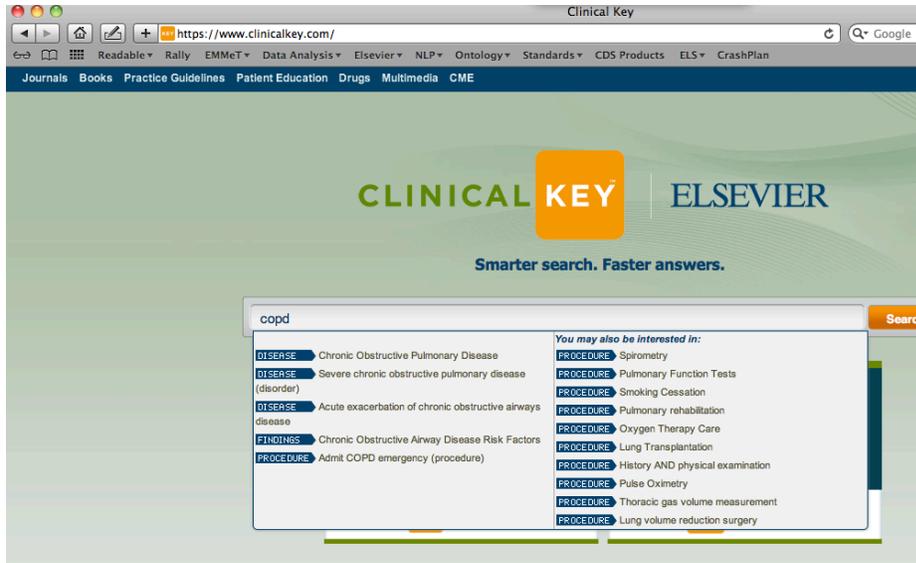
Figure 2: Search text with Auto-complete

to suggest related searches that might be of interest to the user.

The search process uses the indexed content and relevancy scores to retrieve the relevant answers. The results are displayed to the user as a ranked list of titles – journal articles, book chapters, etc. [Figure 3]. The results are also categorized into additional facets such as journals, books, images, videos, etc. and into clinical domain categories (with summary metrics), which can be used for further filtering of results. The preview functionality allows users to quickly review the most highly ranked suggested paragraphs from a given article or book chapter for relevancy before accessing the full text.

A number of additional tools such as the ability to save frequent searches, search results, presentation maker, etc. are also available to users. ClinicalKey launched in April 2012 and is available at:

https://www.clinicalkey.com/

REFERENCES

Bergamaschi S., Guerra F. (2010). *Guest Editors' Introduction: Information Overload.* Internet Computing, IEEE.

Rindflesch, T. C. and Aronson, A.R. (2002). *Semantic Processing for Enhanced Access to Biomedical Knowledge*, in Real World Semantic Web Applications, IOS Press, 157-72.
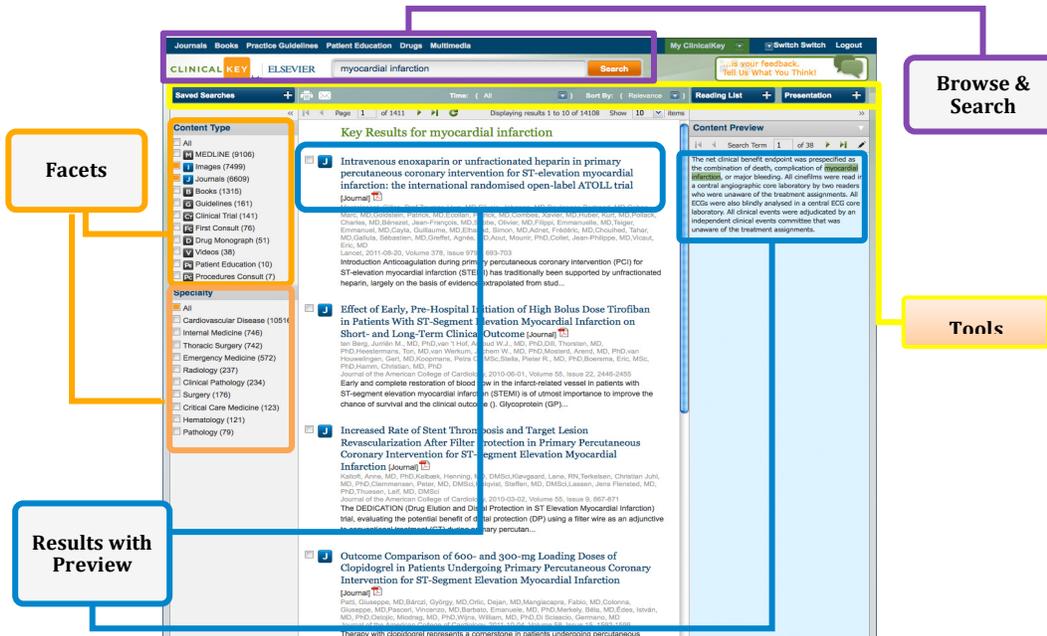
Figure 3: Results with Facets, Preview and Tools