MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

Identifying Anatomical Concepts in Biomedical Text for Automatic Selection of Images

Powell J. Bernhardt[a], Thomas C. Rindflesch[b], Halil Kilicoglu[b], Michele Tringali[c]

[a]*Temple University, Philadelphia, PA, USA*
[b]*National Library of Medicine , Bethesda, MD, USA*
[c]*Azienda Ospedaliera Santa Maria della Misericordia, Udine, Italy*

Introduction

We discuss a method for automatically suggesting appropriate illustrations for anatomically-oriented text. The system uses natural language processing and domain knowledge in the Unified Medical Language System® (UMLS)® and relies on 'part of' information in an anatomical ontology. We discuss the application of our methodology to a set of gastrointestinal endoscopy (GIE) reports. Rather than providing a separate image for each anatomical concept in a report, our system determines the parts of coherent structures being discussed.

Methods

Natural language processing [1] maps input text to anatomy concepts in the Metathesaurus, which are submitted to further processing in the gastroenterology domain. Concepts identified in a GIE report by this processing are highlighted in (1).

1. The endoscope was introduced past the cricopharyngeus down through the **esophagus** and into the **stomach**. **Pylorus** was identified and intubated. Second portion of the **duodenum** and duodenal bulb were examined.

Such concepts are then located in the 'part of' hierarchy in the Foundational Model of Anatomy (FMA) [2] in order to group them into coherent structures that form the basis for suggesting images. A "line of ascent" in the 'part of' hierarchy is computed for each concept. This includes the concept of interest as well as all direct ancestors. For example, the lines of ascent for the anatomical concepts in (1) are shown in (2).

2. [*Duodenum* → Small intestine → Intestine → Gut → Abdomen]
[*Esophagus* → Foregut → Gut → Abdomen]
[*Pylorus* → Stomach → Foregut → Gut → Abdomen]
[*Stomach* → Foregut → Gut → Abdomen]

Lines of ascent are then grouped into "families" sharing a common ancestor, and the lowest common ancestor is computed for each family. For example, the lowest common ancestor for the family in (2) is "Gut."

An image that represents a coherent section of a GIE report should be broad enough to incorporate all relevant terminology, and we suggest that an image of the lowest common ancestor satisfies this requirement. Appropriate illustration should also display anatomical detail. For this, images of the descendants of the lowest common ancestor can be displayed. The user can be given the option of viewing the larger structure (e.g. gut) or detailed images of a specific location or region of that structure (e.g. stomach or esophagus).

Choosing the most useful anatomical images as illustrations for text depends on the needs of the user in addition to the semantic content of the text. Gross anatomical images with labeled structures may be appropriate for students and consumers. However, a specialist may only be interested in gastrointestinal endoscopic images, for example.

Conclusion

We discuss a methodology for suggesting appropriate images to illustrate anatomically-oriented text. Concentrating on gastrointestinal endoscopy reports, we use natural language processing to map text to concepts in the UMLS. The FMA 'part of' hierarchy is used to guide the selection of the best image for a text. The user can then choose a concept to be illustrated and specify a point of view: macro-, micro-, or endoscopic.

References

[1] Sneiderman CA, Rindflesch TC, and Bean CA. Identification of anatomical terminology in medical text. Proceedings of the AMIA Annual Symposium, 1998; pp. 428-32.

[2] Brinkley JF, Rosse C. The Digital Anatomist distributed framework and its application to knowledge-based medical imaging. Journal of the American Medical Informatics Association 1997: 4(3):184-98.

Address for correspondence

Thomas Rindflesch, National Library of Medicine,
8600 Rockville Pike, MS 43, Bethesda, MD 20894, USA.
Email: tcr@nlm.nih.gov. Phone: (301) 435-3191.