Analysis of Medical Student Content Searches that Resulted in Unidentified UMLS Concepts

INTRODUCTION
Many authors have reported on the use of the Unified Medical Language System (UMLS) to match concepts in free text. Unmatched search strings may be due to misspellings, concepts not in the UMLS, or searches for words not expected to be in the UMLS (e.g., names of people or places).

METHODS
We mapped medical student search strings from a full-text, concept-based curriculum database to UMLS concepts using a concept identifier\(^1,2\) and performed a failure analysis. We manually reviewed a random sample of search strings that did not match any UMLS concepts. These were categorized by a medical student and a physician.

RESULTS
Table 1 shows the categorization of searches that resulted in matched concepts from 8/02 - 2/06. Of 28,120 total searches, 24,459 (87%) matched at least one UMLS concept.

Table 2 shows the categorization of the random sample of unmatched search strings. The majority of unmatched searches were medically related (67.9%). Misspellings (51%) and unrecognized abbreviations (14%) were the most common causes of unmatched medically related searches. Only 12% of the missed medical searches were concepts absent in the UMLS.

CONCLUSION
The UMLS effectively represents searches performed by medical students. Misspellings, abbreviations, and partial matches are the most frequent causes of unmatched medical searches; absent UMLS concepts were rare (<2% of all searches). To make searching more powerful, more robust algorithms for linking abbreviations to concepts and correcting spelling errors are needed.

REFERENCES

OBJECTIVE:
To investigate why some searches performed by medical students did not match UMLS concepts.