INTRODUCTION

The Google search engine indexes more than 20 billion distinct web pages, containing millions of pictures and many trillion words of text. Given this resource of almost unimaginable scope and depth, our research group is pursuing methods of leveraging this information to enhance augmentative communication (AAC). To this end, we have developed a web crawler that is capable of autonomously browsing through web pages to collect specific information, based on input from an AAC word prediction program. The information is then processed and used to populate the word prediction dictionary, adding topic relevant vocabulary to the word prediction list.

The purpose of this presentation is to discuss the current status of the web crawler, a project funded Rehabilitation Engineering Research Center on Communication Enhancement, National Institute on Disability and Rehabilitation Research.

Three Web crawler areas of web crawler research and development will be discussed. First, an overview of what the web crawler is and how it operates will be provided. A demonstration of how it is currently being employed to provide topic specific vocabulary for word prediction will be provided. Second, findings from a research project documenting the word prediction performance of the web crawler will be presented. Finally, we will discuss our current research and development efforts in creating an autonomous web crawler that can be used with different word prediction engines.

ACKNOWLEDGEMENTS

The authors wish to acknowledge support from the U.S. Department of Education under grants #H133E980026, #H133E030018, and #RW97076002. The opinions expressed are those of the authors and do not necessarily reflect those of the supporting agency.

The Google search engine indexes more than 20 billion distinct web pages, containing millions of pictures and many trillion words of text.

Finally, we will discuss our current research and development efforts in creating an autonomous web crawler that can be used with different word prediction engines.