

Preface

The striking advance of computers and especially the emergence of Internet ushered in the information society. Information has become wealth of the society as well as a resource of the globe. Today when one wants to acquire any information, say, where to buy an antique camera, which is the most suitable hotel in Birmingham, or some knowledge about Chinese characters and radicals, he can surf the Internet or the World Wide Web to retrieve it. If he gets it, but if it is not in his native language, then he may use a language translation tool which is attached to the search engine. Unfortunately, in many cases he may get into some embarrassing position. For example, the search engine may drown him in hundreds of thousands — if not millions — of results to his retrieval, in which there are so many duplications, and the translation tool may give him a translation so awkward that he may take it as another foreign language. All this shows that it may not be hard to retrieve some information with a keyword on the Internet, but it is not so easy to get precisely what one actually needs; it may not be difficult for a computer to look up a translation of a word in a bilingual dictionary, but it would by no means be easy to translate a real text. When the first computers appeared in 40s of the last century, the scientists' first attempt of its application was to deal with natural languages. However, 60-year experience has proved that natural language processing (NLP) is really a very hard nut to crack.

In order to improve the technology of NLP, the scientists in the field have tried so many ways, one of which is to create large-scale knowledge resources, including machine-tractable dictionaries. HowNet is one of the

cases in which great effort has been devoted to the building of knowledge resources.

HowNet is unique with its definition in a set of basic sememes (the smallest units of meaning) and semantic relations. HowNet has been now widely applied in a lot of subject areas of NLP, such as word similarity measurement, information extraction, text categorization, question answering, word sense disambiguation, etc.

HowNet demonstrates its prominent power in the computing of meanings. If anyone wants to know what the coverage and depth of the meaning computation that HowNet can achieve, he can try the Mini-HowNet and its affiliated tools which are attached to this book. For example, if he wants to know the semantic distance between “doctor” and “dentist”, and between “doctor” and “nurse”, the tool, called Concept Similarity Measure, will tell that the first pair is much closer than that of the second pair. If anyone wants to know how meanings are computed by HowNet, how it can achieve such an extent and depth, or the theory of meaning and its computation, he should read this book. To play with the Mini-HowNet and its affiliated tools while reading this book will be of great fun!

This book is not the termination of the development of HowNet, but the beginning of its new stage. At this moment to look back into the past and recall some of the great events and figures that exercised influence on the development of HowNet will be of significance. In the late 80s, in Makoto Nagao’s frequent academic visits to China, I interpreted for him and learned a great deal from his rich experience in building semantic dictionary, especially the principles for semantic classification of nouns and verbs. From 1987 to 1992 when I was the chief technical leader of Chinese team participating in the machine translation project among Japan and other four Asian countries, I learned quite a lot from various kinds of MT dictionaries of Japanese IT companies and labs, especially from Japan’s EDR concept dictionary. In 1988 I was invited by Antonio Zampolli to attend the Summer School on computational linguistics. During the 1-month course, for the first time I learned ontology from Sergei Nirenburg’s lecture, and visited the knowledge base built in the lab of Pisa University. In 1993 when I worked in Tokyo, I visited Hozumi Tanaka of Tokyo Institute of Technology, during the whole

morning of discussion, he gave me lots of valuable advice and gave me many papers which described taxonomy or discussed semantic categorization. My 3-year work in Institute of Information Systems in Singapore was a substantial preparatory stage for the research of HowNet. I would like to thank all my friends who have rendered their help in the creation of HowNet. At this moment my deep memorial goes to Chen Liwei, late President of Chinese Information Processing Society, who was the respected guide to my research career.

Finally, thanks to my wife who is always a strong supporter of my life and work.

Zhendong Dong