Applying Natural Language Processing Techniques to Bioinformatics

Hyun-Seok Park,

Macrogen Corp.

Abstract

Considering that there is the lack of standards for storing genome-related on-line documents, the techniques in Natural Language Processing (NLP) is likely to become more and more important. It is necessary to extract useful information from the raw text and to store it in a computer-readable database format. Recent advances in NLP technologies raise new challenges and opportunities for tackling genome-related on-line text for information extraction task. For example, we can obtain many useful information related to genetic network or metabolic pathways simply by analyzing verbs such as 'activate' or 'inhibit' in Medline abstracts in a fully automatic way. Thus, combining NLP techniques with genome informatics extends beyond the traditional realms of either technology to a variety of emerging applications.

Curriculum Vitae

Hyun Seok Park, Ph.D.

Macrogen Co.

Seian Building #116, 9F

Shinmun-Ro 1Ka, Chongro-Ku

Seoul, Korea 110-061

Phone: +82-2-3704-4506

FAX: +82-2-370404683

E-mail: hspark@macrogen.co.kr, hsp@sejong.ac.kr

Education and Research Appointments

Bachelor of Science	Seoul National University, Electronic Engineering, Seoul, Korea
Bachelor of Science	University of Kansas, Computer Science, Kansas, USA
Master of Science	University of Pennsylvania, Computer Science, Philadelphia, USA
Doctor of Philosophy	University of Cambridge, Computer Science, Cambridge, UK
Postdoctoral Fellowship	University of Tokyo, Information Science, Tokyo, Japan

Professional Activities

1989, Member of the Golden Key National Honor Society (U. of Kansas, Honor Roll)

1994/5-1994/12, Research Fellow, IRCS Lab, University of Pennsylvania, Philadelphia. USA

1997/10-1998/10, Research Associate, Tsujii Lab, University of Tokyo, Tokyo, Japan

1998/12-1999/2, Visiting Scholar, University of Tokyo, Tokyo, Japan

1998/9-2000/2, Visiting Professor, Sungshin Women's University, Seoul, Korea

1999/7-present, Director, Macrogen Co.

2000/3-present, Associate Professor, Sejong University, Seoul, Korea

Projects Involved

"The Application of Korean-English Machine Translation to a Military Message Domain", DARPA Project, Battelle. Account No. 5-21630; Sponsoring No. 1326 (IRCS Lab, University of Pennsylvania)

"Developing a Korean Version of The Core Language Engine", (SRI & Computer Laboratory of University of Cambridge, Gonville & Caius College)

GENIA Project: Developing an NLP System for Information Extraction Task from Genome Related Documents, JSPS-RFTF96P00502 , (Human Genome Center & TSUJII Lab of University of Tokyo)

Z Project: Whole Genome Random Sequencing and Fragment Assembly of *Zymomonas mobilis*, (Macrogen Co., Seoul, Korea)