# The 90th RCKC Colloquium
## Research Center for Knowledge Communities at University of Tsukuba

第90回知的ｺﾐｭﾆﾃｨ基盤研究ｾﾝﾀｰ 研究談話会

# Title: Cross-Language Entity Linking in 21 Languages

講演者：Prof. Douglas W. Oard
(University of Maryland, USA)

In the traditional view of information retrieval, search engines help the user to find documents, and users then read those documents. For a world in which information is abundant and time is scarce, there are clear limits to the scalability of such an approach. The alternative is to have our machines read documents for us, and then to somehow help us to find and understand what they have learned. This is the perspective that motivates much of the current work on information extraction, knowledge-base population, and linked open data, all of which will be components of some as-yet undesigned system.

In this talk, I will start by briefly reviewing some related projects in the USA, Europe and Japan. I will then focus on one component of such a system that we have been working on at the Johns Hopkins University HLT COE. Our goal is to perform cross-language entity linking, associating mentions of entities that are found in a document written in one language with a knowledge base that was designed originally using some different language.

In this talk, I will focus on a new test collection that we have built in which a mention of a person can be in one of 21 languages and the knowledge base is a 2009 snapshot of English Wikipedia. I'll describe an efficient way to create such a collection using a combination of tools that already exist for English, large collections of parallel text, and some crowdsourcing. We used this approach to create a publicly available multilingual cross-language person-entity linking collection that includes between 875 and over 4,000 queries for each of 21 non-English languages. I will then present some results from our initial experiments with this test collection. I'll conclude the talk with a few forward-looking remarks on the present focus of our knowledge-base population work. This is joint work with Dave Doermann, Dawn Lawrie, Paul McNamee and Jim Mayfield.

2012年7月5日 （木）
July 5, 2012 (Thurs.)
14:00 - 15:00

筑波大学 筑波ｷｬﾝﾊﾟｽ春日ｴﾘｱ
情報ﾒﾃﾞｨｱﾕﾆｵﾝ3階 共同研究会議室1
Lecture room 1 on the 3rd floor of Union of Library and Information-media Studios (ULIS) in Kasuga Area

**言語は英語です。参加費無料，参加申込不要**
**The seminar will be presented in English. No charge to participate and no reservation is needed.**

知的コミュニティ基盤研究センター
Research Center for Knowledge Communities, University of Tsukuba
Tel: 029-859-1524, Ext.:81524
E-mail: kc-office@slis.tsukuba.ac.jp
http://www.kc.tsukuba.ac.jp/colloqium/index.html