NII ESSAY

What is name identification?

Akiko Aizawa
Professor, Digital Content and Media Sciences Research Division, the National Institute of Informatics

All sorts of events occur during the course of peoples' lives—from ordinary things such as births, deaths, marriage and divorce, to adoption and immigration. In all of these cases, various forms of notification have to be made, often dispersed across a wide geographical range. And since these documents are handled by people and therefore liable to human error, misspellings and data input errors can occur.

The difficulty of linking records

Linking each and every one of these geographically scattered individuals' records is somewhat more difficult than it appears, but it is paramount to civil life because it is the basic data for proving a person's identity and providing the proof by which their family allowance and pensions are paid.

Linking records is the process of checking two pieces of data and deciding whether or not the content conforms. Though this would at first appear to be a task that could be performed easily with computers, this is not actually the case. The Von Neumann-type calculating machine was first conceived in 1946, and whilst the computerization of society subsequently progressed rapidly the essential difficulty of the issue of linking records remains unchanged.

Let's take a look at a good example of this difficulty by trying to find out about a certain person using search software. First, we enter the surname - resulting in a mountain of data about the same name and the names of places etc. Entering additional keywords about the person's affiliations etc. should help to narrow down the scope of the search. But in many cases, establishing whether or not the content then displayed pertains to a certain person is a task that eventually has to be done through manual verification. In other words, the role of combing records is also a question of verifying descriptions - whether or not the person in one record is the same as the person in another record.

Verification of statements, and costs

When the role of linking records boils down to the verification of people's statements, one cannot help but be reminded of the "name identification " phrase that filled newspaper columns and became a contemporary buzzword during the recent public pensions problem in Japan, in which millions of payment records were lost.

"I haven't been paid the pension I was supposed to be enrolled in," "Who do the disappearing pensions belong to?" These are simply a question of verifying the veracity of statements about people, and 'identifying their names' to see if they are the same people.

However, this difficulty, which has remained unsolved for over half a century, appears to posses all the essential perplexities of data processing. Vagueness in notation cannot be avoided as long as human intervention is involved, and we have to approach data processing with the premise that this vagueness exists. In doing so, what is important is information about who carried out the verification and in what manner: in other words, the clarification of where the responsibility lies. It is with respect to this responsibility that the considerable costs of verification are entailed.

However, these costs should be considered an investment in the additional merit of clarifying verification and the locus of responsibility, rather than merely as compensating for the losses caused by incomplete information. Therefore, treating unverified statements as if they have actually been verified is liable to incur chaos, which is exactly what we have witnessed in the case of the pensions fiasco.

Furthermore, the fantasy of dramatic cost cuts through automated data processing is, as half a century's research shows, little more than an overestimation of the value of computerization.

The issue of linking records is an excellent example illustrating that the knowledge we automatically obtain from computers and the knowledge verified by experts are both assets, and both essential.

Weaving Information into Knowledge
NII
NII Today No. 23, October 2007 (This English language edition of NII Today corresponds to No. 37 of the Japanese edition)
Published by: National Institute of Informatics, Research Organization of Information and Systems
Address: National Center of Sciences 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
Chief editor: Yoh'ichi Tohkura Cover illustration: Makoto Komori Photography: Shuichi Yuri Design: Kotaro Suzuki Production: Sci-Tech Communications Inc.
Contact: Publicity and Dissemination Team, Planning and Promotion Strategy Department TEL:+81-3-4212-2135 FAX:+81-3-4212-2150 e-mail: kouhou@nii.ac.jp