8th International Conference on Applied Informatics
Eger, Hungary, January 27–30, 2010.

On the Growing Semantic Gap

András Benczúr

Department of Information Systems
Eötvös Loránd University
e-mail: abenczur@inf.elte.hu

Abstract

There is always a semantic gap between a human being and a computer that cannot be eliminated. This is because a large part of the human knowledge is tacit, non formal, hard to express in signointerpratational form. Querying data-systems, retrieving information from computer systems always needs formalized access to the system, and needs a transformation from informal knowledge to formal representation. I call this process filling the semantic gap. In my presentation I will discuss the effect of the growing computerized data systems on filling the semantic gap.

Whether it is true that evolution of info communication technologies will help us to fill the semantic gap, or the growing of the data volume stored on the NET will generate a new kind of growing semantic gap?

Let us consider a simplified model where a person named XY with total knowledge K uses a computer system with total data content M. Both K and M contains the signals or data on thinking or computing. The computer answers a query Q if M contains a program P that computes answer A from Q and M. The conditional Kolmogorov-entropy of A on the condition that M is given – in notational form K(A|M) – is not greater than the length of Q. This is the information non-growth law of the algorithmic information theory.

Suppose then that XY has got a problem D to solve, e.g. to make a decision or action, and knowledge K is not enough to solve D. So XY turns to the computer with formulating a query Q perhaps with some iteration and receives the answer A. Then in this way XY is able to obtain a solution for S using K extended with A. Similarly to the answer A of the computer, we have to estimate K(S|K,A), the conditional algorithmic information quantity of S given K and A. This leads to difficulties because K and reasoning process in K is not formal. The quantitative analyses of this simplified model demonstrate that the size of Q and A must increase with the growing volume of M. And a new problem arises when the size of Q and A exceeds the „processing and storing capacity" of XY. This may lead to a new and growing semantic gap.

Keywords: knowledge, information retrieval, large databases, algorithmic information quantities

MSC: 68P15, 68P30