As hardware and software limitations become less and less rigid, users tend to maintain and use larger and larger databases. Unfortunately, sooner or later they run into a problem that databases become too large to grasp by a human user, hence useless in practice.

Conventional tools and techniques provide adequate means for the acquisition, processing and reporting of data in the traditional sense. However, for the human being the only natural communication means is natural language which, in spite of its inherent imprecision, does provide useful information. Moreover, natural language can best reflect the very essence (in an extremely condensed form) of information.

Fuzzy logic has been advocated for this purpose for some time, and Yager’s (1982) [cf. Kacprzyk and Yager (2000)] approach may be viewed as one of most relevant proposals; we will also employ its idea here. Basically, Yager derives linguistic database summaries as linguistically quantified propositions, e.g., “most of the employees are young and well paid”, with a degree of validity (truth, ...), in case of a personnel database.

This may be viewed as a manifestation of the computing with words paradigm [cf. Zadeh and Kacprzyk (1999a, b)], and we present a new approach to such a linguistic summarization explicitly based on that paradigm.

First, since the number of attributes is usually very high, a fully automatic derivation of summaries is not realistic because the system does not know by itself what (class of) summaries the user may be interested in. Instead, we advocate some human assistance [cf. Kacprzyk and Zadrozny (2000)]. First, we assume that the user specifies the subject of the summary and the attributes of interest. Moreover, the user can also specify more sophisticated structures of summaries as, e.g., involving fuzzy relations. Basically, these are equivalent to the formulation of a database query with a fuzzy linguistic quantifier - cf. Kacprzyk and Zadrony’s (1997, 2000) FQUERY for Access, a fuzzy logic based add on to Microsoft Access. A recent extension of FQUERY for Access that incorporates elements of Bosc and Pivert’s (1992, 1995) SQLf extension of SQL is mentioned and shown that it may increase the efficiency of the summarization process [cf. Kacprzyk and Zadrony (2001)].

Having defined the interesting class of a linguistic summary, we define for each possible instantiation of the summary from that class its degree of quality [cf. Kacprzyk and Yager (2000)], and find the best summary.
We present some results of implementation of the proposed method for the derivation of linguistic summaries of a database concerning sales in a medium size computer retailer in Southern Poland [cf. Kacprzyk and Strykowski (1999)]. The summaries obtained have been very positively accepted by the owner, and have given much valuable insight into diverse aspects of planning and day to day activities.

Notice that the above data summaries are derived on the basis of database contents. However, there also exist some “external” data available from distributed databases and mostly accessible through the Internet. These contain, for instance, data from climatic and weather databases, financial databases, etc. These pieces of data may provide very relevant tools for deriving qualitatively different, more general and relevant types of linguistic summaries, called here linguistic metasummaries that basically summarize own data within a broader context of outside data.

We present a formal problem statement for the derivation of such linguistic metasummaries, and then consider basic issues related to their implementation. In particular, we discuss the implementation through the technology of intelligent agents. We show how both free and commercial external sets of data can be used. We also mention how to deal with problems with data sets that contain, e.g., contradictory information.

We show the derivation of such metasummaries for a computer retailer when, additionally, temperature and precipitation information (available from the Internet, using intelligent agents) is employed.

Clearly, the approach proposed may find many uses in all kinds of human-centered computer systems, notably in e-commerce, decision support, etc.

Bibliography


