Formalisatie van informatieplanning

Kiewiet, D.J.

SUMMARY.

Information architecture specification.

Nowadays many organizations pay attention to Information Planning. Several definitions exist of Information Planning, but one common aspect in all these definitions is the attention for the future information systems support of an organization. So, Information Planning leads to the definition of an information systems plan for the organization. The specification of all the information systems together in such an Information Plan is the information system architecture of the organization.

The starting point for the specification of such an information architecture is formed by the information processing activities in the organization, the so-called business processes. These business processes operate on entity types: aggregates of data in the organization. Usually, there are three possibilities for business processes to operate on entity types: 1) a business process only uses data of a certain type, 2) a business process creates or changes data of a certain entity type, and 3) a business process does nothing with the data of a certain entity type. Business processes, entity types and operations are represented in an information matrix, with the rows representing the business processes, the columns representing the entity types, and the cells representing the types of operation (a "U" in a cell denotes possibility 1: use, a "C" denotes possibility 2: create, and a "0" denotes possibility 3: nothing).

For getting an information architecture, business processes and entity types in an information matrix are grouped in several information areas, which are the bases for information systems. This grouping process is not trivial: research has shown that experienced information planners were not able to judge information system architectures in a consistent and consequent manner. In order to make it possible that information planners have computer-support for the specification of an information architecture, a formalization of information architecture specification is necessary. However, such a formalization does not exist yet. At the end of chapter 1, this leads to the following research-objective and -questions:

- objective:

  formalizing information architecture specification;

- questions:

  1. *to what extent is it possible to formalize the appreciation of an information architecture;*
  2. *to what extent is the consideration of information architecture specification as a classification problem helpful in generating information architectures.*

In Chapter 2 traditional methods of information architecture specification are described and criticized. One commonly used family of methods for information architecture specification transforms the information matrix so that the columns of the information matrix are transposed so that the 'C-cells' of the matrix are on the topleft-bottomright diagonal of the matrix. Around this diagonal rectangles are distinguished, where each rectangle identifies an information area. In these processes of transforming and distinguishing, three objectives are considered:

- localization of data-creation;
- minimization of data-exchange;
- maximization of coherence within information systems;

One major criticism of these information architecture specification methods is that these objectives are *not* operationalized. Because of this, information architecture specification has a substantive subjective component. One of the conclusions at the end of Chapter 2 is, that more objective descriptions for information architecture specification have to be looked for.

The Generalized Classification Model.

In Chapter 3, operationalization of the objectives mentioned before was performed by making a mathematical model of information architecture specification: the Generalized Classification Model. Two sets are the bases for the model, one set (B) representing the business processes in the organization for which the information architecture has to be specified, the other set (D) representing the entity types. On the Cartesian product of these two sets a mapping u is defined, representing the possible operations between business processes and entity types.

Subsets of B × D represent information areas, and in order to define an information architecture, the notion of a block-partition is introduced (notation: P(B × D)).

In the Generalized Classification Model, operationalization of localization of data-creation has been done by looking in the information matrix to the neighbours of a cell in the information matrix in which a "C" is placed. Quantification was realized by using the function ME which takes in account, for each pair from B × D, the values on the mapping u for the neighbours of that pair. How larger the value on ME, how more localization of data-creation there is. However, the value on ME is dependent on a certain ordering of the elements of B and D. A larger value on ME, i.e. more localization, is perhaps possible if another ordering of the elements is looked at. Therefore, those orderings (or permutations) of the elements of B and the elements of D have to been looked for, that maximizes the function ME.

In order to operationalize the objective of minimization of data-exchange, data-exchange was defined by introducing two relations $\mathfrak{R}_c$ and $\mathfrak{R}_p$ in the Generalized Classification Model. The relation $\mathfrak{R}_c$ is used to define data-exchange between two pairs of B × D, and this relation is used to specify the relation $\mathfrak{R}_p$ which defines data-exchange between two blocks of P(B × D). Quantification of data-exchange is in the model achieved by introducing two functions: the function S_1 called pair-cohesion, and the function S_2 called block-cohesion. The function S_1 quantifies the presence of an $\mathfrak{R}_c$ relation. The function S_2 denotes for two blocks the number of couples of pairs, one pair in each block, for which there exists a $\mathfrak{R}_c$ relation. The total data-exchange for a block-partition can now be defined as the sum of S_2-values for all possible couples of different blocks in the block-partition.

Coherence within an information system can be operationalized as the amount of data-exchange between two pairs belonging to one block, each pair representing a business process and an entity type. When using this operationalization, the third objective for information architecture specification can be quantified in the model by using the functions S_1 and S_2.

Subsets of B × D represent information areas, and in order to define an information architecture, the notion of a block-partition is introduced (notation: P(B × D)).

In the Generalized Classification Model, operationalization of localization of data-creation has been done by looking in the information matrix to the neighbours of a cell in the information matrix in which a "C" is placed. Quantification was realized by using the function ME which takes in account, for each pair from B × D, the values on the mapping u for the neighbours of that pair. How larger the value on ME, how more localization of data-creation there is. However, the value on ME is dependent on a certain ordering of the elements of B and D. A larger value on ME, i.e. more localization, is perhaps possible if another ordering of the elements is looked at. Therefore, those orderings (or permutations) of the elements of B and the elements of D have to been looked for, that maximizes the function ME.

In order to operationalize the objective of minimization of data-exchange, data-exchange was defined by introducing two relations $\mathfrak{R}_c$ and $\mathfrak{R}_p$ in the Generalized Classification Model. The relation $\mathfrak{R}_c$ is used to define data-exchange between two pairs of B × D, and this relation is used to specify the relation $\mathfrak{R}_p$ which defines data-exchange between two blocks of P(B × D). Quantification of data-exchange is in the model achieved by introducing two functions: the function S_1 called pair-cohesion, and the function S_2 called block-cohesion. The function S_1 quantifies the presence of an $\mathfrak{R}_c$ relation. The function S_2 denotes for two blocks the number of couples of pairs, one pair in each block, for which there exists a $\mathfrak{R}_c$ relation. The total data-exchange for a block-partition can now be defined as the sum of S_2-values for all possible couples of different blocks in the block-partition.

Coherence within an information system can be operationalized as the amount of data-exchange between two pairs belonging to one block, each pair representing a business process and an entity type. When using this operationalization, the third objective for information architecture specification can be quantified in the model by using the functions S_1 and S_2.

Minimization of data-exchange between blocks and maximization of coherence are combined into one objective function F.

At the end of Chapter 3 it is argued that the problem of finding those permutations of B and D which yield an optimal value for ME (a subproblem of the maximization of F), is NP-hard. So, for obtaining a solution, one has to rely on heuristics.

Cell manufacturing formation.

In Chapter 4 another interpretation of the Generalized Classification Model is given. This interpretation is the problem of cell manufacturing formation. In this problem, machines and parts (both needed for producing a certain product) are arranged into cells, making use of routing information. This routing information describes which parts are processed on which machines. The cells are so constructed that machines that process the same parts, and parts that are processed by the same machines, are grouped into one machine-part workcell. So, information architecture specification and cell manufacturing formation resemble each other: they are interpretations of the same Generalized Classification Model. The benefit of this similarity is that several heuristics, used often in the area of cell manufacturing formation, can be applied in the area of information architecture specification.

Classification techniques and information architecture specification.

Three heuristics are chosen to exploit the similarities between information architecture specification and cell manufacturing formation. These three heuristics are cluster analysis, mixed integer programming, and neural networks. In Chapter 5 these heuristics are described, and it is shown formally that these heuristics can be applied for solving the Generalized Classification Model, and thereby information architecture specification.

In Chapter 6 these three heuristics are compared to each other on the performance of the objective function F of the Generalized Classification Model. This comparison is done by using the three heuristics for specifying information architectures for a realistic case. This case is the CBR (Central Bureau for certificates of driving proficiency), the national institution in The

Netherlands responsible for administering exams and for providing medical statements, both with respect to driving proficiency.

Due to the size of the problem (51 business processes and 30 entity types), it was not possible to specify an information architecture with help of the heuristic Mixed Integer Programming. On the other hand, the use of cluster analysis and neural net gave seven information architectures. The best information architecture with respect to the F-value of the Generalized Classification Model is the one obtained by using the neural net.

In Chapter 7 it is concluded that the research-objective (formalization of information architecture specification) is attained. In the future, this formalization can be used in the development of computer-support for information planners, and with that for the monitoring of the information planning process.