Josef Zelger
Universität Innsbruck, Institut für Philosophie

A Dialogic Networking Approach to Information Retrieval

Prof.Dr. Josef Zelger (b.1940), studied theology, physics, philosophy and psychology. Full professor at Innsbruck University, author of: Konzepte zur Messung der Macht, 1975; Konflikte und Ziele, 1986; Verfahren kreativer Selbstorganisation, 1988. Co-founder and co-editor of the journal "Conceptus" since 1967.

Zelger, J.: **A dialogic networking approach to information retrieval.**
Knowl.Org. 21(1994)No.1, p.24-28, 5 refs.
The user of documents, e.g. of a library, faces the task of exploring the data fundus and of selecting information according to his actual intentions. He even may find or happen to find new aspects and, following these, will further develop his original quest. How can the user be supported by PC-based procedures? Recently at the Institute of Philosophy, University of Innsbruck/Austria, the GABEK method (*Ganzheitliche Bewältigung sprachlich erfaßter Komplexität*) was developed. It has proved useful so far in similar cases of ordering and/or retrieving information; especially to build hidden order structures and to incorporate them into information processing and storage facilities. It seems that the GABEK method might be applied successfully also to the user problem as mentioned above. To clarify his quest the user relies on a database$_1$. This base contains experiences of previous users, which are expressed in natural language sentences. Through a PC-supported dialogue$_1$, founded on database$_1$, the user elaborates a more detailed concept of his own topic. This concept later is termed a "linguistic gestalt", if it fulfils certain conditions. The linguistic gestalt may include 3 to 10 sentences in natural language, which specify the user's original intentions. The key terms contained in this linguistic gestalt will, in a dialogue$_2$, be employed to retrieve relevant information from database$_2$. Database$_2$ represents the information system, e.g. a library. The procedures as indicated above and the building of linguistic gestalts can be effected by GABEK. Small quantities of data provided, the WINRELAN program (1993) may be used.
(Author)

1. The GABEK Method

Before describing our method of information retrieval we must give an idea of the GABEK method. In principle the method works by representing fields of knowledge, intent, attachment, etc. by means of conceptual networks. Even unsorted arguments, ideas, texts, notes, quotations can be presented in the form of a comprehensive, clearly arranged "map". Here conceptual relations are first traced between unordered contents. By means of the conceptual networks one can move from one content in the network to the next one. The focal structures of the network, moreover, are determined by key concepts as chosen by the user. Arbitrarily he can focus his attention on one concept, image or sentence and then move to the next one. Statements are connected within the network through at least one shared concept. GABEK represents a method of putting much-quoted networked thinking into practice.

It is assumed that many users of the information system have been asked e.g.: "What are you searching for?" "What is your field?" "What are your expectations, problems and assumptions?", etc. In the GABEK procedure the variety of user opinions can be used to draw holistic pictures of actual research problems in special areas, joining up the many partial aspects, i.e. bringing them into a meaningful network. This can be done by PC-support.

The PC-program WINRELAN provides file cards on which the text can be entered. Every answer is written on a separate file card. Then the user marks on each file card those expressions he identifies as key concepts. Thus the possible nodes in the network are determined. In the computer the sentence A (A: *I am working on the encouragement of weaker pupils through team teaching in elementary schools.*) is saved as text and as relation between the following expressions.

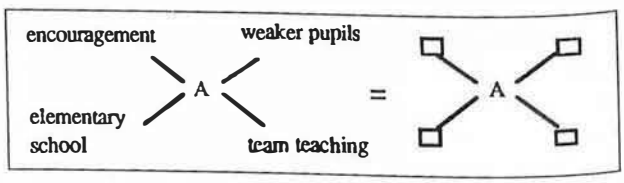

Figure 1: The sentence A presented as a relation between four expressions

In figure 2 one can see that the answer B (B: *In my experience team teaching leads in small classes to more strain on weaker pupils.*) is related to the answer A via two shared concepts.

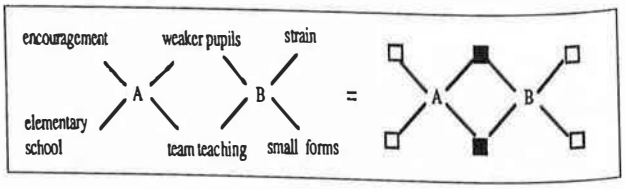

Figure 2: Sentences related via 2 nodal expressions

When longer texts are presented in this way, we obtain a very complex conceptual network. occurring in more than one statement; white ones for individually occurring expressions.

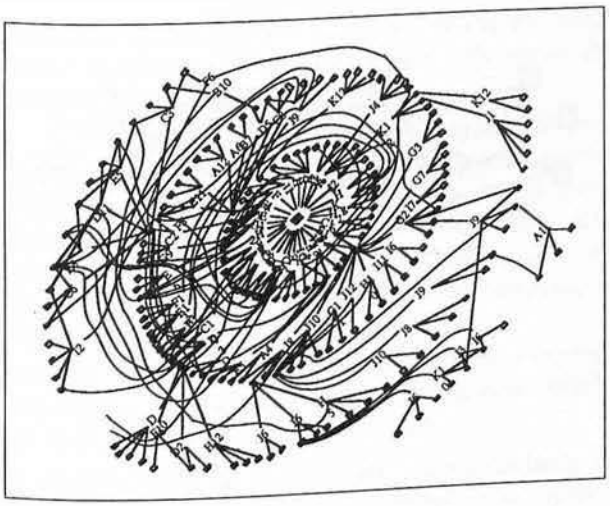

Figure 3: A conceptual network

How can we analyse the conceptual network or resulting partial networks? We intend to support a procedure which we know from everyday thought processes: we focus our attention on one concept, one image, or one perception. But we cannot stop there. We are dealing with cognitive objects wich we are unable to call to mind in isolation. When we become conscious of them, these are related to further cognitive objects and these relations are stabilized through shifts of attention like movements of a beam of light. Strictly speaking, an isolated thought which is not related to other ones is cognitively impossible, i.e. only an abstraction, as we know that every thought brings another one with it. We can neither retain a single sense impression without perceiving other sense data nor can we focus on a single concept without moving on to further concepts or to stop at an image without moving on to other ones. Nevertheless, these are units which do not continually merge into each other. A method for analysing complex contents should therefore accommodate the concentration on a restricted field of attention as well as the shift of attention along semantic relations.

GABEK (4) represents a pragmatically applicable method encouraging holistic, networked thinking. This is supported by WINRELAN (2), enabling interactive monitor work and permitting step by step decisions on which paths to follow in the opinion network, which aspects of evaluation are to be pursued, which contents are to be focussed on and which are to be ignored. The user explores the meaning relations like autoroutes. He always has to decide which of the varied opinions (landscapes) interest him most. The result of an analysis or synthesis is always a selection of highlighted interrelated statements, which are stored in the form of the original expressions used.

There are some different formal methods for analyzing the expression network (4). Here it is to be shown how the user of the database can develop and clarify his own goals or his questions to be answered through the documents.

2. GABEK as a Heuristic for Problem Definition in Form of a Linguistic Gestalt

2.1 Interacting with Database$_1$

Suppose we have answers of many users to the question "What are you searching for?", etc. They have been ordered according the above method with the help of the PC-Program WINRELAN (2). The resulting conceptual network of experiences, i.e. the database$_1$ is now offered as an assistance tool to users.

Intially in dialogue1 the user will have in mind little but a general idea of his topic and intention. He will undertake to clarify and to specify them. Dialogically he asks the database$_1$ for related information and its context, its internal and external relationships, etc. He will use the answers received as an input to develop the formulation of his problems to be solved and his aims as intended.

Thus he begins to organize his own network of knowledge or of questions. To this effect he accesses the problem descriptions and the experiences of previous users stored in the PC$_1$. These are stored and organized by the program in full sentences. By freely navigating through them, following his own judgment, the user gradually shapes the formulation of his topic as he thinks appropriate. The final sentences will be added to database1 by the PC$_1$. Thus the outcome of the procedure is twofold: a final formulation of the user's question and an improved PC-stored network of user's knowledge.

2.2 Organization of the Selected Sentences in Form of a Linguistic Gestalt

According to his interests the user selects first one or two concepts and reads the sentences in the network of database$_1$, which contains these concepts. If he does not accept these sentences as valuable for his research he then continues searching for other statements in the database$_1$ or he writes down new ones. Continuing in this way, he tries to formulate about 3 to 10 sentences which express his research problems.

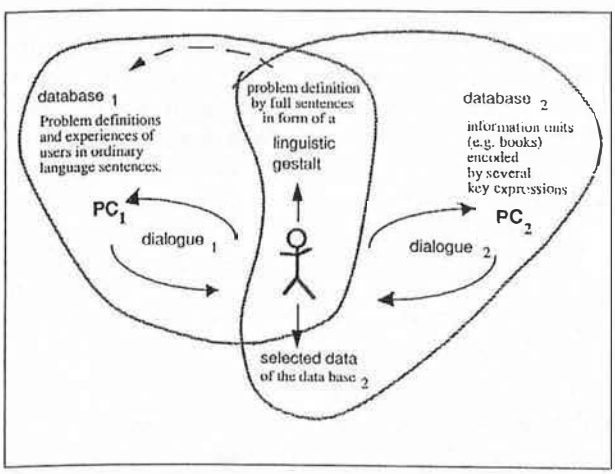

Figure 4: Two main phases of dialogue can be distinguished. Of course, both phases may be recursively employed.

The sentences selected should of course further the user's understanding. I presuppose that we don't learn through isolated sentences but by meaningful chunks or clusters of sentences. These must be consistent and coherent. Therefore I introduce the concept of a "linguistic gestalt". We learn to understand linguistic utterances as coherent linguistic complexes which are comprehended as linguistic gestalts. Such meaningful wholes are to be built actively by the individual in a self-organized way using the conceptua$_1$ network of database$_1$, presented as object of learning (5).

A linguistic gestalt is an abstract entity. Like perceptive gestalten it presupposes grouping of the intentional object. These parts are statements or simply relations between concepts. Viewing them pragmatically I call them "units of meaning". The linguistic gestalt can be distinguished pragmatically from the whole linguistic context through interrelation of units of meaning within the linguistic gestalt. Thus we interpret a linguistic gestalt as a specific linguistic network which can be seen from a syntactic, semantic or pragmatic perspective.

Viewed syntactically, a formal linguistic gestalt consists of relations between groups of expressions within an expression network.

I call these groups of expressions "sentences".

A *formal linguistic gestalt* G_{synt} is a finite subset of the set of sentences (= sets of words) in the expression network S: $G_{synt} \subset S$, and G_{synt} has to fulfil the subsequent conditions:

Formal variety (internal distinction of the parts, differentiation):

In G_{synt} no expression may occur at more than one node (elimination of duplicate expressions) No sentence (= set of expressions) in G_{synt} may be included in another sentence of G_{synt}. And G_{synt} may only contain sentences which are distinguished from each other by more than one key expression.

Formal connectivity

At least three expressions of each sentence S in G_{synt} must be integrated within G_{synt} in a strictly circular form. That is, for each of at least three expressions of each sentence S in G_{synt} there must exist a path to at least two other sentences in G_{synt} and vice versa, where the two paths differ. This presupposes that these expressions are nodal expressions.

Formal distance

The necessary steps to reach from each sentence S_i in G_{synt} to each other sentence S_j in G_{synt} must not exceed the maximal value of 2 steps.

For illustration we show a formal structure which fulfils the rules. The dark squares represent different nodal expressions; $S_1, \ldots, S_{10}$ indicate ten different sentences. The structure is a result of a selective analysis of open-ended interviews. All the conditions are fulfilled in the structure (Fig.5), where even all expressions are interrelated in a strictly circular way.

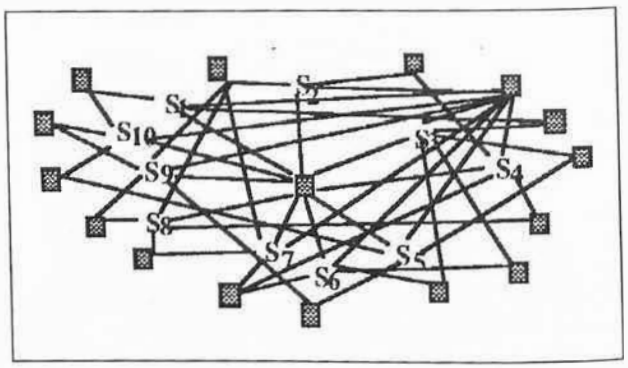

Figure 5: A formal linguistic gestalt

Analyzing texts, we observe that it is not easy to comply with the above conditions. They oppose each other and lead to a kind of equilibrium when they are fulfilled. The more the sentences are internally differentiated, the more likely they are not sufficiently networked internally. And the closer the network, the more likely it is that internal differentiation does not suffice.

From a semantic point of view the semantic linguistic gestalt is a net of semantically determinative relations between concepts, a conceptual network.

The *conceptual network* is defined as an expression network, so that all sentences So in the expression network are interpreted statements or meaning postulates between concepts. These semantically determinative relations must be demonstrable intersubjectively by means of an ideal (paradigmatic) example. (I.e. it must be possible to point out to an instance, i.e. an ideal, paradigmatic example for the semantically determinative relations.)

From a pragmatic perspective the linguistic gestalt is a network of units of meaning for an individual x in the situation s and at the point in time t, i.e. an opinion network.

A unit of meaning is seen as the connection of subjective concepts. It is considered as subjectively meaningful by an individual x in a situation s at a point in time t.

The *opinion network* is defined as conceptual network, so that all statements or the semantically determinative relations between concepts in the conceptual network are also units of meaning for x in s and t. There must exist an individual x, a situation s and a point in time t for which it is true that x believes in the situation s and at the point in time t that all units of meaning of the conceptual network are applicable to perception, orientation or action patterns.

2.3 Linguistic Gestalt-formation by Dialog$_1$ with PC-Support

Subsequently an interactive method for constructing linguistic gestalten is to be shown . The PC-program WINRELAN developed for GABEK allows the fulfilment of the above proposed conditions for linguistic gestalten. *The PC controls the formal conditions. The*

fulfilment of the semantic conditions and of the pragmatic adequacy is judged by the user. We sketch the method very briefly:

a) If we have unsorted texts like the results of open-ended interviews, then we start the subprogram cluster analysis. The texts appear in a machine-sorted form. According to a preselected level, the answers are sorted in groups so that all texts in a group are connected strongly through common expressions. Many of the clusters therefore fulfil the rule of formal connectivity.

b) The user selects one cluster and within it some statements according to semantic and pragmatic criteria or he writes down new ones by himself.

c) By redundance analysis the PC shows pairs of sentences of the selected ones (b) which are not distinguished enough from each other, for they do not fulfil the rule of formal variety.

d) The user reads the statements which are not distinct enough and *decides which ones are to be deleted or combined and edited in a new guise.*

e) Through coherence analysis the PC controls whether single expressions infringing against the rule of formal connectivity occur.

f) The user reads the statements and *decides whether they should be shortened* (by eliminating concepts which occur only once (fa) *or whether there should be added new statements* connecting concepts occurring only once in a denser and wider conceptual network (fb). In the first case (fa) the user reduces the statements to ensure that only repeatedly recurring key concepts play a significant role and returns to step (c). In the second case (fb) *through goal-oriented navigation he attempts to complete the incomplete network* by further statements. This is achieved by not remaining within the proposed statements but by proceeding from the concepts ocurring only once within the chosen group of sentences to navigate across the statements in the linguistic network, relating the concepts ocurring only once to other concepts.

g) The PC *suggests texts, or sentences which are connected through common concepts* with the ones proposed earlier.

h) The user reads the statements and decides whether he will accept them as a linguistic gestalt. If this is not the case he continues the goal-oriented navigation.

This procedure is continued until the user can accept the sentences as a linguistic gestalt representing his aims and fulfilling the pragmatic, semantic and syntactic conditions. This "*subjective selection, going hand in hand with an unconscious modification of the search goal, cannot be made by anyone else but himself.*" (Recommendations for Knowledge Organization (1, p.152).

3. Information Retrieval from Database$_2$ Using the Linguistic Gestalt

Now we still need information about the stored documents in the database$_2$. With his topic in form of a fully developed linguistic gestalt the user enters dialogue$_2$, employing the now well formed key expressions as a foundation from which to search for related units of information and to select them. The well established structural base laid down in dialogue1 enables him to

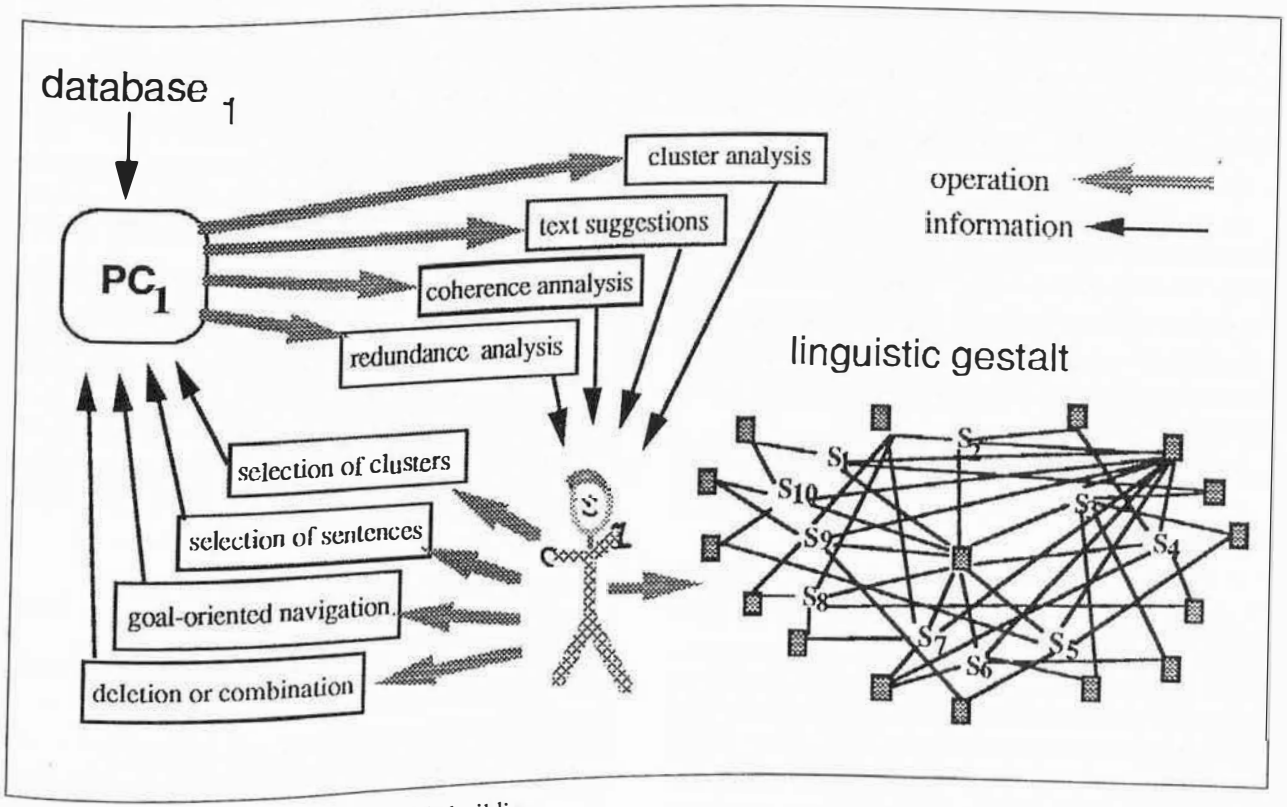

Figure 6: Procedure of linguistic gestalt-building

search and - if he wants to, provisionally in a hermeneutic mode - to incorporate new information rather freely without too much danger of loosing the original aim or context. In addition, the process has gained self-monitoring qualities. For the search procedure the user will rely on the linguistic gestalt which is the controlling base of the information retrieval process.

The user asks which documents in database$_2$ may correspond to the nodal concepts of his linguistic gestalt. We suppose that each document of database$_2$ is described by some key concepts. These concepts form units in the information system (for each unit there exists a document). If one key concept is used to describe more documents, then this is a nodal concept in the conceptual network of the information system.

The dialog$_2$ now is running as follows:

a) The PC$_2$ selects all documents which are encoded by at least one of the nodal concepts of the linguistic gestalt. The resulting list may be too extensive for practical use because one of the problems is always connected with the "noise" of information retrieval. "*A dozen useful responses among thousands of other ones are practically undiscoverable. Often an information system supplying useful œnoiseÆ and thus initially functioning to the full satisfaction of its users needed to be abandoned after having gradually or suddenly (e.g. when nobody is available any more to help weed out the noise) proven to be largely useless.*" (Recommendations for Knowledge organization (1, p.151)). So we need to restrict the responses - but in a intelligent way. In our procedure the intelligence is given through the coherent organization of the linguistic gestalt. Thus we will restrict the responses:

b) The user selects one sentence which is most relevant according to his aims. He identifies the set of nodal expressions of this sentence (e.g. the nodal concepts of the sentence S_1) and searches for documents which are encoded by the set of the nodal concepts of S_1.

c) The PC$_2$ shows the corresponding documents.

d) If only few documents are found, then the user reduces the number of expressions (of the same sentence) which are to be met by the documents.

If too many documents are shown then the user selects a second sentence (e.g. S_2) related with S1 and he extends the number of expressions which should be met by the selected documents.

d) The PC$_2$ shows the corresponding list of documents.

c) The user selects documents from the list according to his purposes.

f) The user goes on to further sentences S_3 .... to S_{10} in his linguistic gestalt and looks for further documents which fulfil the corresponding conditions.

In this way the user can control the number of documents or his own according to his linguistic gestalt. Thus the gestalt principles secure a basic but flexible order of documents found and the continuity of the knowledge ordering or learning process of the user.

4. Problems

The opposite problem to the one mentioned is the problem of too much information loss. It can be reduced if we include more units (documents) in the conceptual network selected from database$_2$. We could for instance include all units which are connected immediately in database$_2$ with one unit found above (see above a).

The problem of too much information loss is of course also connected to synonyms. We can treat this problem in the following way: The user will meet some expressions in the conceptual network of database$_2$ which occur in his linguistic gestalt and others which are not contained. He can use the first ones like starting points in his retrieval procedure. For some of the others he must decide whether he will judge them as synonyms with expressions in the linguistic gestalt. If he does so, then this decision of the user is stored in the concept network of database$_2$. Earlier proposed synonyms are deleted. For new research processes the last proposed synonyms can also be used by newcomers in the concept network of the database$_2$. In this way the information system is learning from the users without losing the order given by the original pre-established documentation vocabulary.

Of course it would be necessary to test the proposed method empirically. The GABEK method has proved useful in other applications. But the problems arising from practical implementation in the field of information retrieval are not yet known.

References

(1) International Society for Knowledge Organization, Working Group Content Analysis: User Evaluation of Information Systems. Int. Classif. 19(1992)No.3, p.151-152 = Recommendations for Knowledge Organization, R-01, Oct. 1992.
(2) WINRELAN 1993 - developed for GABEK by J. Schoenegger & J. Zelger, Institut für Philosophie der Universität Innsbruck, Innrain 52, A-6020 Innsbruck.
(3) Zelger, J.: A Holistic Method of Mastering Complexity. In: Klein, H. (Ed.): Managing Change with Cases, Simulations, Games and other Interactive Methods. Needham, Boston: WACRA 1991. p.255-267.
(4) Zelger, J.: GABEK, a New Method for Qualitative Evaluation of Interviews and Model Construction with PC-Support. In: Stuhler,E., Suilleabhain, M.O. (Eds.): Enhancing Human Capacity to Solve Ecological and Socio-economic Problems. München, Mering: Rainer Hampp Verlag 1993. p.128-172.
(5) Zelger, J.: Von sprachlichen Gestalten zur Hypergestalt. In: Philosophie und Verfahren kreativer Selbstorganisation. Innsbruck: Institut für Philosophie der Universität Innsbruck 1993. Preprint Nr. 26.

Notes:

Paper presented at the First Russian ISKO Conference, Moscow, May 10-14, 1993.
I wish to thank to Dr. Philip Herdina and Dr. Hellmut Löckenhoff for suggestions and criticism.

Prof.Dr. Josef Zelger, Leopold Franzens Universität
Institut für Philosophie, Innrain 52, A-6020 Innsbruck, Austria