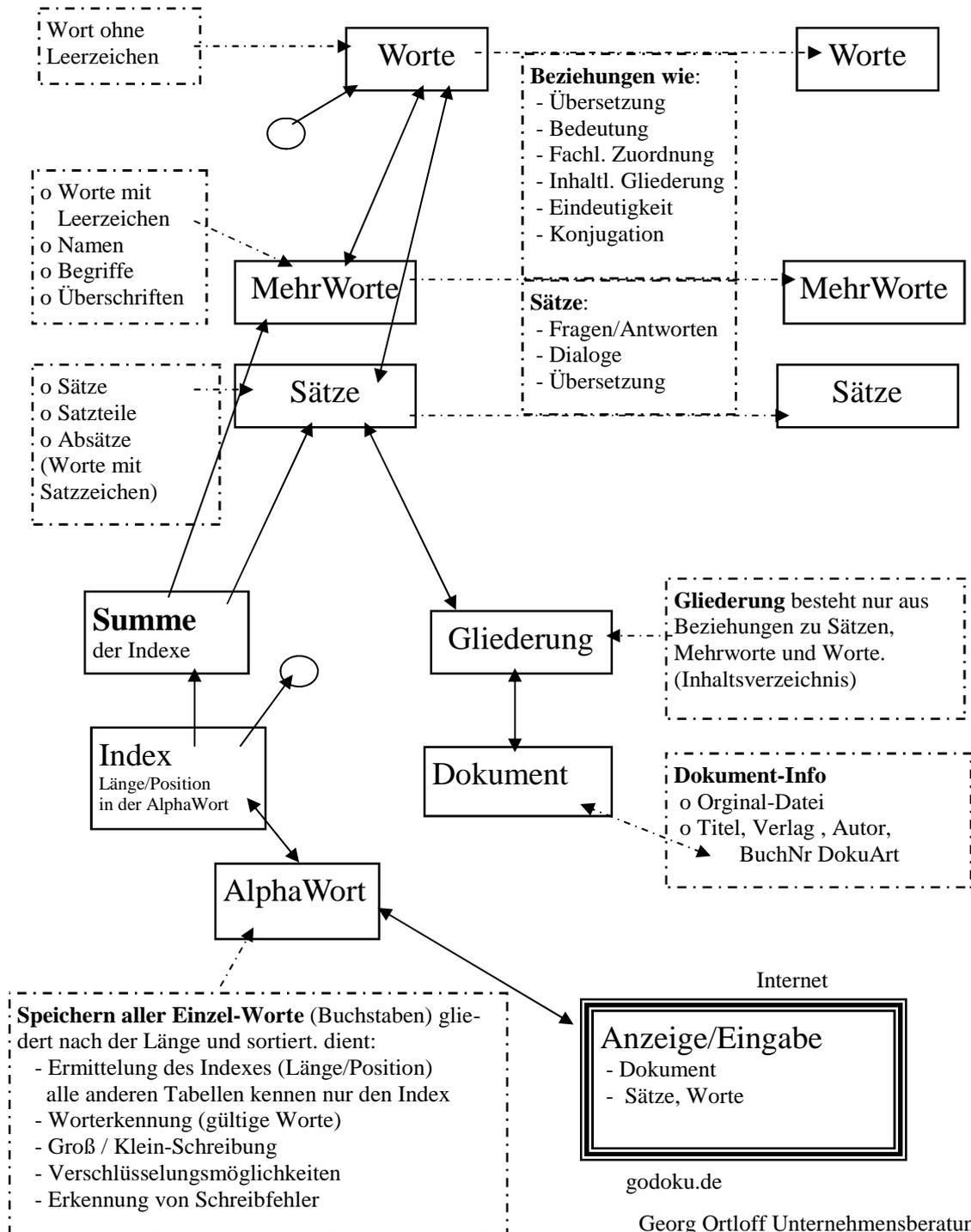


# Was ist das Doku-System

# Wissensdatenbank

Es ist ein Datenbanksystem in dem „grenzenlos“ Worte und deren Beziehungen gespeichert werden können.  
 Es ist keine Verwaltung von Dateien und deren Inhalte.  
 Es ist eine System, das folgende Beziehungen verwaltet:



# Wissensdatenbank

## Speicherung der Information

### Aufbau des Beziehungsteiles

	Pos 1	.. 2	3—10	..11	..12	..13
Wort	Beugung	Geschlecht	z.Z. 8 Positionen für eine Inhaltliche Gliederung  wird über eigene Tabelle definiert	Beziehungen zu		
MehrWort	Worte (Index)	sort. Worte		Doku	Mehr	Satz
Satz	Worte (Index)					

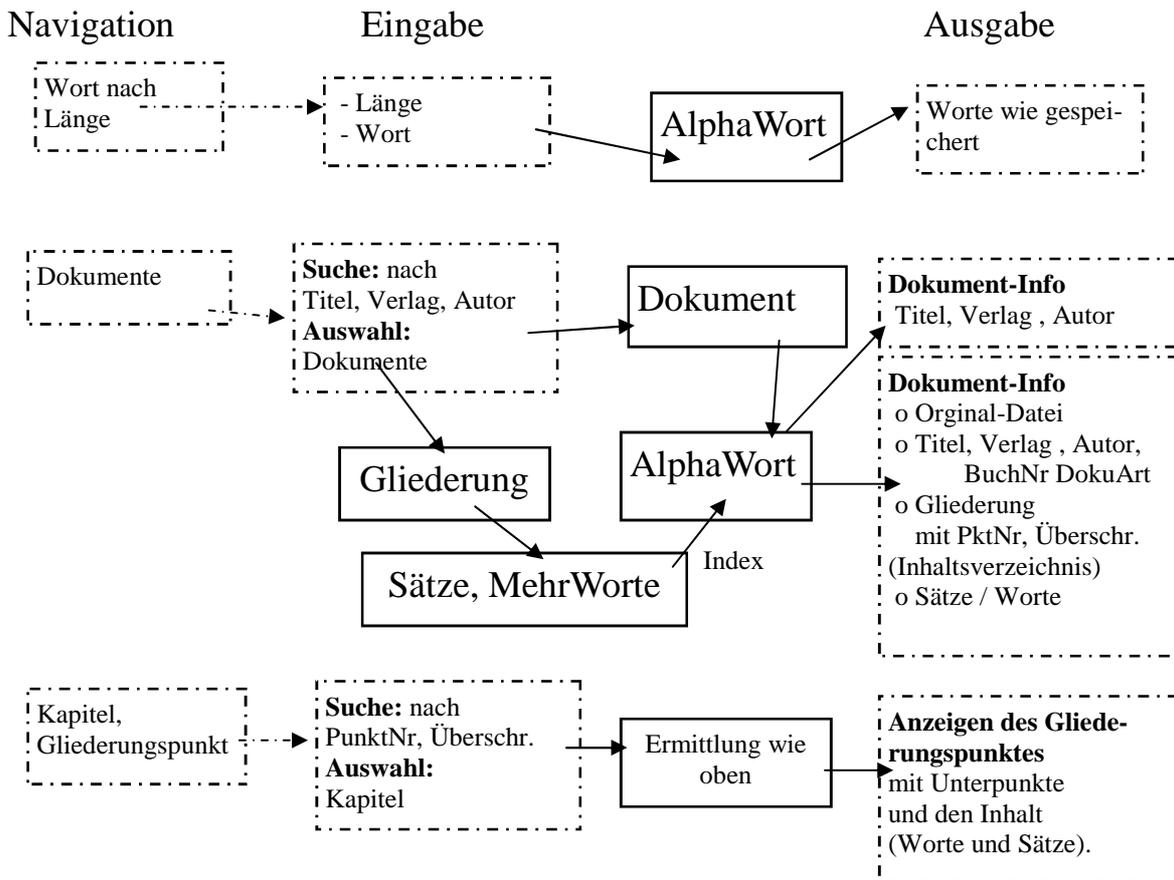
### Aufbau der Doku-Tabellen

Wort	Index (binär)	Trenng	Grund- Wort	Sprache Key	Status (vollst.)	fachl. Gliedg	inhaltl. Gliedg	Bezieh. (s.oben)
MehrWort	lfd.Nr	Summe Indizes	Grund- Wort	Sprache Key	Status (vollst.)	fachl. Gliedg	inhaltl. Gliedg	Bezieh. (s.oben)
Satz	lfd.Nr	Summe Indizes	Anz. Worte	Sprache Key	Status (vollst.)	fachl. Gliedg	inhaltl. Gliedg	Bezieh. (s.oben)
Gliederung	lfd.Nr	Key Doku	Anz. Worte	Sprache Key	Gliedg Punkt	Gliedg. Überschr.	Gliedg Stufe	Sätze (Vektor)
Dokument	lfd.Nr	Titel	Verlag	Autor	BuchNr	DokuArt	Sprache	Beschreib.

	Sprachbeziehungen			
Wort	Key	Anz.	Sprachen	Beziehungen
MehrWort	je Tabelle	Übersetz.	Vektor	Vektor
Satz				

## Wie funktioniert das Doku-System Wissensdatenbank

Zentrale Bedeutung für das Dokusystem hat die **AlphaWort-DB** in ihr sind alle bekannte Worte nach Länge und aufsteigend sortiert. (**direkte Suche**). Gleichzeitig ist die Position auch der Index für die Speicherung in den anderen DB's.  
 Da das Dokument nicht als Dateien gespeichert ist, wird der Inhalte komplett über die DBs: **Dokument, Gliederung, Sätze, MehrWorte und AlphaWort** neu erstellt.



**Inhalt der DB's:** **Dokument** -> Titel, Verlag , Autor, BuchNr DokuArt  
**Gliederung** -> PktNr und Überschrift sowie die Keys der Sätze und MehrWorte  
**Sätze / MehrWorte** -> aus den Indizes der Worte über die **AlphaWort** -> wird das Wort zugeordnet.

Testen Sie wie performat dies geschieht.

Durch den strukturierten Aufbau kann der einzelne Satz und auch das einzelne Wort mit allen seinen Beziehungen angezeigt werden.

**These:**  
 wenn dieser Dokumentenaufbau so performat ist, dann ist Kommunikation mit dem Computer möglich: d.h. Fragen —> Antwort ; Fehler —> Problemlösung ;  
 Aufbau von Dialogen ; usw.

## Wie funktioniert das Doku-System Wissensdatenbank

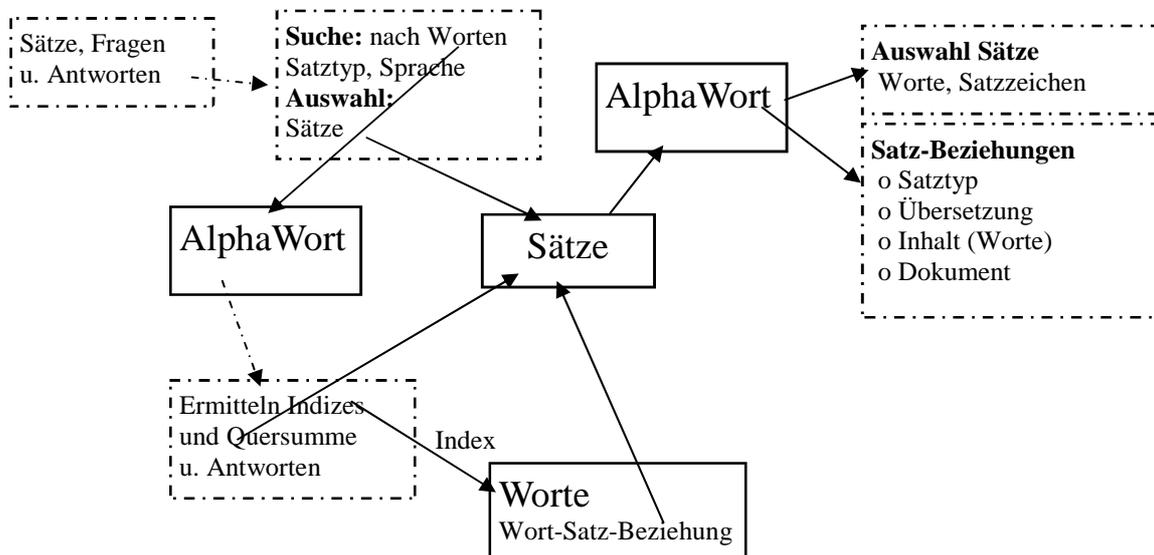
Durch die Gliederung der Dokumente in einzelne Sätze entsteht unabhängig von den Dokumenten die Möglichkeit zusätzlich Beziehungen zwischen den Sätzen herzustellen. z.B. Übersetzungen auf Satzebene, Fragen —> Antworten

**Jeder Satz im Doku-System (Worte und Satzzeichen) ist nur einmal vorhanden**, dies ist durch die **direkte Satzsuche (Quersumme)** möglich. d.h. durch die Indizes (Worte) ist eine Summierung der Worte (Quersumme) möglich. Weiterhin kann über die **Wort-Satz-Beziehung** sehr schnell Sätze gefunden werden. d.h. der Satz mit den meisten gewünschten Worten wird zu erst angezeigt.

Navigation

Eingabe

Ausgabe



Durch die **Einmaligkeit des Satzes** (nur einmal vorhanden auch wenn dieser Satz in mehreren Dokumenten vorkommt) macht es Sinn diesen inhaltlich zu gliedern. Hier kann je nach Satztyp unterschiedliche Gliederung vorgenommen werden.

z.B bei Fragen —> Antworten zuordnen ; satzgerechte Übersetzung

Hier bestehen auch programmtechnisch viele Möglichkeiten den Satz in seine Teile zugliedern.

Weiterhin kann auch dem Satz eine Bedeutung zugeordnet werden, ist diese eindeutig, so ist eine maschinelle Reaktion möglich. Dies kann ein Prozess oder Antwort sein.

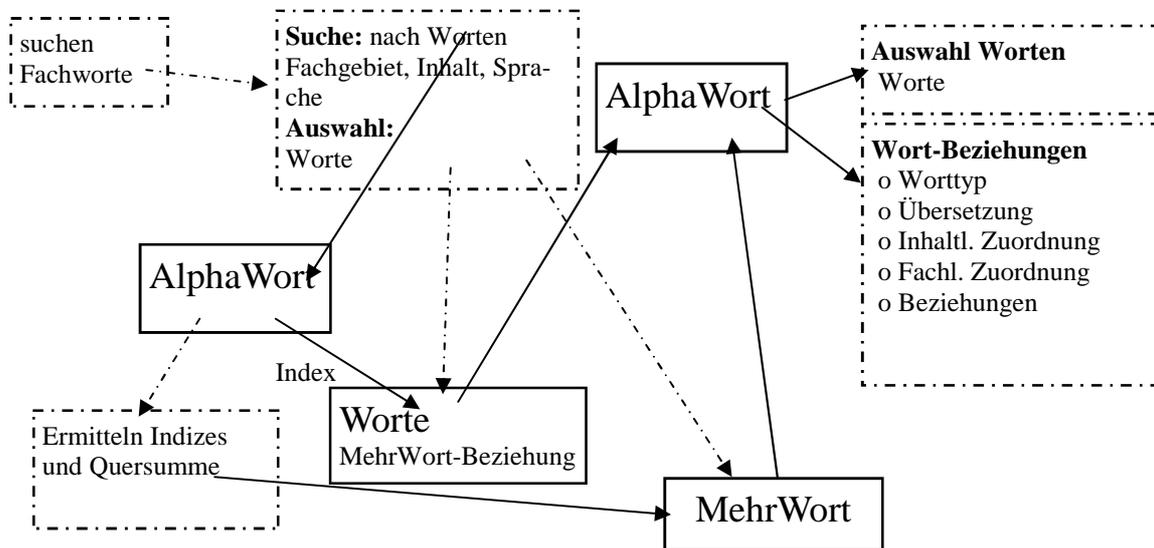
## Wie funktioniert das Doku-System Wissensdatenbank

Die kleinste Einheit ist das Wort (ohne Leerzeichen) dieses wird in der AlphaWort-DB in Buchstaben gespeichert. Diese DB dient auch zur direkten Suche von Worten.  
 Besteht das Fachwort aus zwei oder mehreren Worten (MehrWort), so kann dieses Mehrwort auch über die Quersumme (wie Satz) direkt gesucht werden.  
 Für jedes Wort kann eine fachliche sowie inhaltliche Zuordnung, ggf. auch Worttyp, erfolgen, sowie Sprache, Status und Worttrennung.

Navigation

Eingabe

Ausgabe



Jedes **Wortes** ist nur einmal vorhanden und steht allen Dokumenten, Kapiteln und Sätzen zur Verfügung, deshalb ist es sehr effektiv und wirkungsvoll Wortbeziehungen zupflegen.

An erster Stelle steht eine sinnvolle Zuordnung (Fachlich, Inhaltlich ggf. Worttyp) der Worte

**Folgende Beziehungen sind geplant:**

- o Übersetzungen in beliebig viele Sprachen.
- o Konjugation, Deklination, Artikel
- o Wort durch Worterweiterung —> die Mehrdeutigkeit auflösen (auch für Übersetzung nötig)
- o Thesaurus, Wortherkunft und sinnverwandte Wörter
- o Festlegen der Wortbedeutung

**Vorrangige Aufgabe** ist die Erarbeitung einer **sattelfesten Gliederung** (Zuordnung und Inhalt) um sinnvoll den Datenbestand aufzufüllen.

Wenn ich **Ihre Interesse** geweckt habe und Sie bereit sind mich **zu Unterstützung** (Fachlich, Datenbeschaffung, Finanziell) bin **sehr dankbar**.

Ja **mein Herz** hängt daran, weil ich überzeugt bin, etwas **sinnvolles** zutun.

## Technische Daten

## Wissensdatenbank

Das Datenbanksystem im GODOKU ist nicht nur „Performat“ sondern ist auch von der Größe der Datenbanken scheinbar „grenzenlos“ erweiterbar. Zur Zeit belegt das Doku-System etwa **40 MB Speicherplatz**, sodass es interessant ist, wie sich bei Erweiterung der Einträge (Dokumente, Sätze und Fachworte) sich der Speicherplatz entwickelt.

	Anzahl Einträge	Datenbank- Größe	
AlphaWort	133.000	1,4 MB	<b>Speichern aller Einzel-Worte</b> (Buchstaben) gliedert nach der Länge und sortiert. dient: - Ermittlung des Indexes (Länge/Position) - Worterkennung (gültige Worte) - Groß / Klein-Schreibung - Verschlüsselungsmöglichkeiten - Erkennung von Schreibfehler
Worte	133.000 incl. Beziehung zum Satz	16,4 MB	<b>Beziehungen wie:</b> - Übersetzung - Bedeutung - Fachl. Zuordnung - Inhaltl. Gliederung - Eindeutigkeit - Konjugation
MehrWorte	14.000 incl. Beziehung zum Satz	2,0 MB	
Sätze	57.000 mit 4.500.000 Worte und Satzzeichen	18,4 MB	<b>Sätze:</b> - Fragen/Antworten - Dialoge - Übersetzung o Sätze o Satzteile o Absätze (Worte mit Satzzeichen)
Gliederung	11.000	1,5 MB	<b>Gliederung</b> besteht nur aus Beziehungen zu Sätzen, Mehrworte und Worte. (Inhaltsverzeichnis)
Dokument	87	0,1 MB	<b>Dokument-Info</b> o Orginal-Datei o Titel, Verlag , Autor, BuchNr DokuArt