

Universität Ulm
Abteilung Künstliche Intelligenz

ExtrAns

**Verarbeitung natürlicher,
schriftlicher Sprache**



Überblick

- Motivation
 - Einleitung
 - Eigenschaften natürlicher Sprache
- Überblick ExtrAns
 - Linguistische Module
 - Semantischer Interpreter
 - Benutzerfragen
 - Antwort-Extraktion
- Logische Formen
- ExtrAns Merkmale
- Fragen / Diskussion



Motivation

- Technische Dokumente als wichtige Wissensquellen
- Umgang mit technischen Dokumenten schwierig.
- Beispiel: Man page von screen über 3800 Zeilen.

- Erkennung natürlicher Sprache sinnvoll.
- Bedeutung statt oberflächliche Form der Sprache.

Einleitung

- Natürliche Sprache in Sätze gegliedert.
- System muss Sätze verstehen können.
- Zuerst: Wichtigste Bestandteile von Sätzen erkennen.
 - Objekte (Nomen)
 - Ereignisse (Verben)
 - Eigenschaften (Adjektive, Adverbien)
- Andere Bestandteile (z. B. Artikel) werden durch Link Grammar herausgefiltert.



Eigenschaften natürlicher Sprache

- **Unregelmäßigkeiten und Besonderheiten in natürlicher Sprache erkennen.**
- Schreibfehler
 - Automatische Erkennung wie bei Textverarbeitungsprogrammen.
 - Manuelle Erkennung anhand WorldNet Datenbank.

Eigenschaften natürlicher Sprache

- **Unregelmäßigkeiten und Besonderheiten in natürlicher Sprache erkennen.**
- Hyponyme (Wörter in untergeordneter Beziehung zu anderen Wörtern)
 - Essen ist ein Hyponym von zu sich nehmen.
 - Tablette ist ein Hyponym von Medikament.
- Manuelle Erkennung, Datenbank erstellen.

Eigenschaften natürlicher Sprache

- **Unregelmäßigkeiten und Besonderheiten in natürlicher Sprache erkennen.**
- Homonyme (Gleiches Wort, verschiedene Bedeutungen)
 - Schloss (Türschloss und Gebäude)
 - Bank (Möbel und Geldinstitut)
- Manuelle Erkennung, Datenbank erstellen.

Eigenschaften natürlicher Sprache

- **Unregelmäßigkeiten und Besonderheiten in natürlicher Sprache erkennen.**
- Synonyme (Verschiedene Wörter, gleiche Bedeutungen)
 - Fehler und Defekt
 - Horizontale Position und waagerechte Position
- Teilweise automatische Erkennung durch Fastr.

Eigenschaften natürlicher Sprache

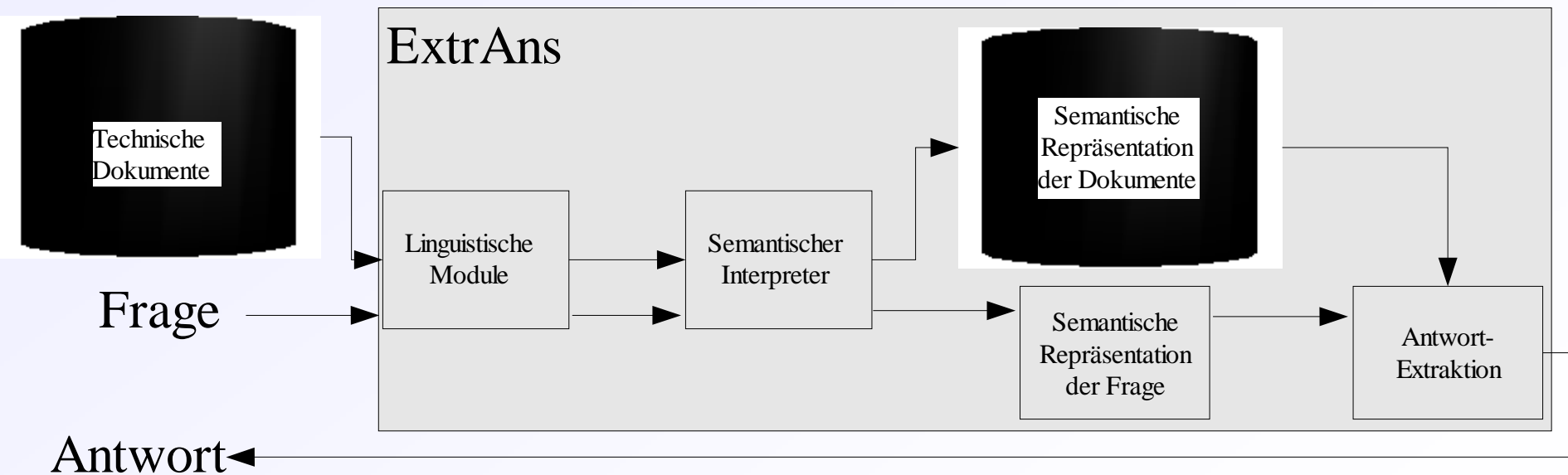
- **Unregelmäßigkeiten und Besonderheiten in natürlicher Sprache erkennen.**
- Erkennung muss vor weiterer Verarbeitung durch ExtrAns geschehen.
- Linguistische Module werden semantischem Interpreter vorangestellt.



Zusammenfassung

- **Natürliche Sprache schwer in Formeln beschreibbar.**
- **Unregelmäßigkeiten und Besonderheiten erfordern intelligente Programme**
- **Universelle Datenbanken müssen erstellt werden.**

Überblick ExtrAns



Linguistische Module

- Unwichtige Wörter werden von Link Grammar herausgefiltert.
- Erkennung Linguistischer Eigenschaften:
 - Synonymie (Fehler und Defekt)
 - Disambiguierung (Auflösung von Mehrdeutigkeiten)
 - Anapherauflösung (Wiederholung eines Wortes zu Beginn aufeinander folgender Sätze oder Satzteile)

Linguistische Module

- Erkennung Linguistischer Eigenschaften (Fortsetzung):
 - Einfache Inversion (“generator control unit” und “control unit for the generator”).
 - Morphologische Variationen (“electric contactor” und “electrical contactor”).
 - Komplexe morphosyntaktische Variationen (“electrical generation equipment” und “equipment for generating electricity”).
 - Kombinationen der oben aufgeführten Variationen (“functional test” und “operational check”).

Semantischer Interpreter

- Generiert logisches Abbild, sog. logische Form pro Satz.
- Logische Form und Zeiger zu Quelltext wird in Wissensbasis gespeichert.
- Semantische Repräsentation der Dokumente wird erstellt.
- Prolog Notation



Benutzerfragen

- Werden in ähnlicher Weise wie Quelltext behandelt.
- Durchgehen linguistische Module
- Semantischer Interpreter erstellt semantische Repräsentation der Frage.



Antwort-Extraktion

- Suche nach passenden Sätzen durch Unifikation in der Wissensbasis zu der semantischen Repräsentation der Frage.
- Antworten werden nach Relevanz sortiert.
- Übersichtliche Ausgabe an Benutzer

Beispiel

The screenshot shows a window titled "WebExtrAns: Main window" with a menu bar containing "file" and "options". The "Data:" field shows the path "/home/ludwig/rinaldi/WEBEXTRANS/datafiles/airbus/". A search query "How is the distribution network supplied?" is entered in a text box, with "OK" and "Clear" buttons to its right. Below the query, the results are displayed as a list of entries, each with a score and a description. The descriptions are highlighted in red. At the bottom, a progress bar shows the search progress from 0% to 100%.

file options Data: /home/ludwig/rinaldi/WEBEXTRANS/datafiles/airbus/

Query (all): How is the distribution network supplied? OK Clear

Ratio of success for the query: 1

Score: [-1,1.000,0]
sqb1k.24.22.00.00/DESCRIPTION/3: In normal flight configuration, each IDG supplies its own distribution network via its Line Contactor (GLC).

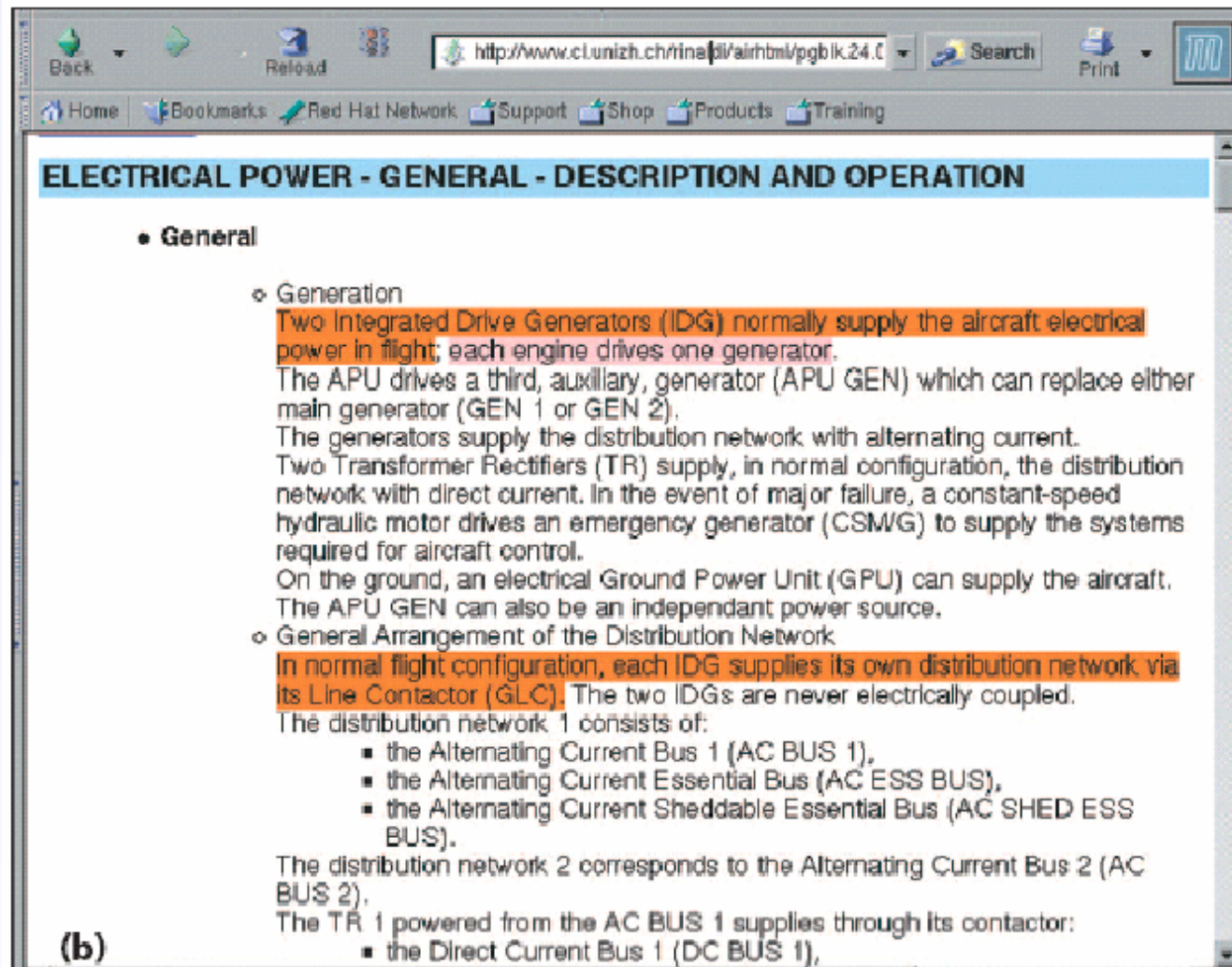
Score: [-1,1.161,0]
sqb1k.24.22.00.00/DESCRIPTION/11: When the two engines run in normal conditions, generator 1 and generator 2 supply their own network .

Score: [-1,1.167,0]
sqb1k.24.22.00.00/DESCRIPTION/238: Then , the generator supplies its network .

Score: [-1,1.400,0]
sqb1k.24.22.00.00/DESCRIPTION/12: Two Integrated Drive Generators (IDG) normally supply the aircraft electrical power in flight; each engine drives one generator supplies network

(a) 0% 25% 50% 75% 100%

Beispiel



The screenshot shows a web browser window with the address bar containing the URL <http://www.c.lunizh.ch/rina/pl/airhtml/pgblk.24.c>. The browser's navigation bar includes buttons for Back, Reload, Search, and Print. Below the navigation bar, there are several menu items: Home, Bookmarks, Red Hat Network, Support, Shop, Products, and Training. The main content area has a blue header with the text "ELECTRICAL POWER - GENERAL - DESCRIPTION AND OPERATION".

- **General**
 - **Generation**

Two Integrated Drive Generators (IDG) normally supply the aircraft electrical power in flight; each engine drives one generator. The APU drives a third, auxiliary, generator (APU GEN) which can replace either main generator (GEN 1 or GEN 2). The generators supply the distribution network with alternating current. Two Transformer Rectifiers (TR) supply, in normal configuration, the distribution network with direct current. In the event of major failure, a constant-speed hydraulic motor drives an emergency generator (CSMWG) to supply the systems required for aircraft control. On the ground, an electrical Ground Power Unit (GPU) can supply the aircraft. The APU GEN can also be an independant power source.
 - **General Arrangement of the Distribution Network**

In normal flight configuration, each IDG supplies its own distribution network via its Line Contactor (GLC). The two IDGs are never electrically coupled. The distribution network 1 consists of:
 - the Alternating Current Bus 1 (AC BUS 1),
 - the Alternating Current Essential Bus (AC ESS BUS),
 - the Alternating Current Sheddable Essential Bus (AC SHED ESS BUS).

The distribution network 2 corresponds to the Alternating Current Bus 2 (AC BUS 2). The TR 1 powered from the AC BUS 1 supplies through its contactor:

 - the Direct Current Bus 1 (DC BUS 1),

(b)

Zusammenfassung

- Linguistische Module bereiten Quelltext für Weiterverarbeitung vor.
- Semantischer Interpreter generiert dann logische Form.
- Benutzerfragen werden in ähnlicher Weise wie Quelltext behandelt.
- Antwort-Extraktion besteht aus:
 - Unifikation von logischer Form der Sätze der Wissensbasis und des Fragesatzes.
 - Geordnete Ausgabe.



Logische Formen

- Repräsentieren Quellsätze
- Anforderungen:
 - Einfach zu erstellen
 - Leicht zu verwenden

Logische Formen

- Flache Notation erfüllt Anforderungen
- Beispiel:
- Geschachtelte, logische Form
 - **A := (factorial(25) - exp(12) * 2);**
- Flache, logische Form
 - **factorial(X, 25);**
 - **exp(Y, 12);**
 - **A := (X - Y) * 2;**
- Neue Argumente X und Y.



Logische Formen

- Flache Notation erfüllt Anforderungen
- Neue Argumente X und Y.

Logische Formen

- Flache, logische Ausdrücke
- Beispiel: „The ECAM contactor is located in the left frame.“
 - **holds(e4),**
 - **object('ecam_contactor',o2,[x2]),**
 - **evt(locate,e4,[x4,x2]),**
 - **object(anonymous_object,o4,[x4]),**
 - **object(frame,o7,[x7]),**
 - **prop(left,p2,[x7]),**
 - **prop(in,p5,[e4,x7]).**

Logische Formen

- Minimale, logische Formen
- Enthalten keine überflüssigen Informationen.
- Beispiel:
- Frage: "Where is the ECAM contactor located?"
 - **object('ecam_contactor',01,Y),**
 - **evt(locate,E2,[X,Y])**

Logische Formen

- Antwort: „The ECAM contactor is located in the left frame.“
 - `holds(e4),`
 - `object('ecam_contactor',o2,[x2]),`
 - `evt(locate,e4,[x4,x2]),`
 - `object(anonymous_object,o4,[x4]),`
 - `object(frame,o7,[x7]),`
 - `prop(left,p2,[x7]),`
 - `prop(in,p5,[e4,x7]).`

Logische Formen

- `object('ecam_contactor',01,Y),`
 - `evt(locate,E2,[X,Y])`
- Durch Unifikation wird Antwort auf Frage gefunden.
- `holds(e4),`
 - `object('ecam_contactor',o2,[x2]),`
 - `evt(locate,e4,[x4,x2]),`
 - `object(anonymous_object,o4,[x4]),`
 - `object(frame,o7,[x7]),`
 - `prop(left,p2,[x7]),`
 - `prop(in,p5,[e4,x7]).`

Zusammenfassung

- **Logische Formen sollen Quellsätze repräsentieren.**
- **Flache Notation**
 - Einfach zu erstellen
 - Leicht zu verwenden
- **Durch Unifikation wird Antwort auf Frage gefunden.**

ExtrAns Merkmale

- Vorteile gegenüber einfachen Stichwort-Suchmaschinen:
 - Terminologie bestimmter Fachgebiete wird berücksichtigt
 - Komplexe und fehlerhafte Sätze können verarbeitet werden.
 - Wissensbasis enthält nur relevante Informationen da flache, logische Form.



Fragen / Diskussion