

Natural Language Processing

Aus Fließtext wird eine auswertbare Datenzeile

Der größte Teil dessen, was ein Unternehmen weiß, steht nicht in der Datenbank, sondern in E-Mails, PDFs, Chats und Verträgen. Für ein IT-System ist das zunächst eine bedeutungslose Zeichenfolge. NLP macht daraus Felder, mit denen sich rechnen und steuern lässt. Der Nutzen entsteht dabei nicht durch das Werkzeug, sondern erst durch den Anschluss an einen Prozess.

DER WEG VOM TEXT ZUR DATENZEILE

- 1 **Unstrukturierter Text.** Kundenmail, gescannte Rechnung, Support-Chat, Vertrag.
- 2 **Vorverarbeitung.** Sätze und Wörter erkennen, Formatierungsreste entfernen, Sprache und Fachbegriffe bestimmen.
- 3 **Aus Beispielen lernen.** Hunderte markierte Rechnungen zeigen dem Modell, wo der Betrag steht. Danach findet es ihn auch in einem unbekannten Layout.
- 4 **Klassifizieren und auslesen.** Der Text bekommt eine Kategorie, und einzelne Werte werden herausgezogen. Zusammen ergibt das einen strukturierten Datensatz.
- 5 **In den Prozess geben.** Weiterleiten, mit der Bestellung abgleichen, bei Fristen alarmieren, Stimmungen im Dashboard zeigen.

NLP UND NLU

NLP
Der Überbegriff. Macht Texte für Computer lesbar und verarbeitbar.

NLU
Ein Teilbereich davon. Fragt nach der Bedeutung: Welche Absicht steckt dahinter, welche Stimmung, welcher Zusammenhang?

DIE ZWEI ARBEITSFORMEN

KLASSIFIKATION
Ordnet einen Text einer Kategorie zu: Reklamation oder Bestellung, dringend oder nicht.

EXTRAKTION
Zieht einzelne Werte heraus: Betrag, Währung, Kundennummer, Frist, Ansprechpartner.

WO ES SICH IM BETRIEB RECHNET

Bereich	Was ausgelesen wird	Was danach passiert
Buchhaltung	Rechnungsnummer, Beträge, Steuersatz, IBAN, Datum	Buchungssatz landet im ERP, weniger Tippfehler
Verträge	Laufzeit, Kündigungsfrist, Haftung, Sonderrechte	Vorsortierung nach Risiko, ersetzt keine juristische Prüfung
Kundenservice	Thema, Dringlichkeit, Stimmung	Ticket geht automatisch ans richtige Team
Produkt und Markt	Wiederkehrende Themen aus tausenden Rückmeldungen	Trends statt Einzelmeinungen im Dashboard

SO TESTET MAN ES AN EINEM NACHMITTAG

- 1 **Einen Fall abgrenzen.** Nicht „unsere Dokumente“, sondern „Eingangsrechnungen“.
- 2 **Zehn echte Beispiele nehmen.** Aus dem Posteingang, nicht aus einer Vorlage.
- 3 **Strukturiert anfordern und gegenprüfen.** Felder und Kategorien vorgeben. Was zuverlässig erfasst wird und was nicht, beantwortet die Frage nach dem eigenen System.

EIN PROMPT, DER DAS ZEIGT

VORGABE STATT WUNSCH
„Lies aus dem Rechnungstext diese Felder aus und gib sie als Tabelle zurück: Lieferant, Rechnungsnummer, Datum, Betrag, Währung, Zahlungsziel, IBAN. Fehlt ein Feld, schreib leer statt zu raten.“
Ohne feste Feldliste kommt Fließtext zurück, und der ist nicht weiterverwendbar.

MERKSÄTZE

NLP macht Texte lesbar, NLU erfasst ihre Bedeutung.

Der Einstieg braucht kein IT-Großprojekt: ein Anwendungsfall, ein paar Beispieltexte, ein Werkzeug.

Ergebnis strukturiert anfordern und Kategorien vorgeben, sonst ist es nicht weiterverwendbar.