

Trendbriefing: Automated Journalism

März 2017

MEDIALAB



Automatisierte Texterstellung: Warum ein Thema?

- ▶ **Personalisierung: Inhalte können individualisiert werden**

In Zukunft können durch die Kombination von strukturierter Erfassung mit Nachrichten- und Kundendaten immer besser auf den Nutzer zugeschnittene Einzeltexte generiert werden. Auf diesem Weg lassen sich bald individualisierte Nachrichten- oder E-Commerce-Seite erstellen.

- ▶ **Versionierung: Eine Ausgangsbasis - viele Texte**

NLG-Software ist in der Lage, auf Knopfdruck viele verschiedene Versionen eines Textes zu erstellen. Damit kann das Dilemma von „Duplicated Content“ umgangen werden, der von den Suchmaschinen abgestraft wird. Auch die Tonalität kann für das jeweilige Medium bzw. demografische Gruppen optimiert werden. Und sogar unterschiedliche Plattformen (Web, Mobile) eines Mediums könnten so mit passenden Angeboten bedient werden.

- ▶ **Long Tail: Angebote für die Nische**

Derzeit ist es aus Kostengründen oft nicht wirtschaftlich, thematische oder lokale Nischen zu bedienen. Im Sportbereich werden im Fußball etwa nur die großen Ligen gecovered, obwohl die Daten für eine Vielzahl an Sportarten vorhanden wären. NLG ermöglicht hier eine gut aufbereitete Berichterstattung, die User-Interessen in weit größerem Ausmaß als bisher bedienen kann.

- ▶ **Schnelligkeit: Texte im Sekundentakt**

Texte können innerhalb von wenigen Sekunden erzeugt und verschickt werden.

- ▶ **Internationalisierung: Sprachbarrieren überwinden**

Anbieter von NLG-Software bieten ihren Kunden die Möglichkeit, ihre Inhalte in mehreren Sprachen zu erzeugen. Das Computerprogramm generiert die jeweiligen Texte direkt in der gewünschten Zielsprache ohne den Umweg über eine andere Sprache. Kostspielige Übersetzungen können so entfallen.

- ▶ **Selbstoptimierende Inhalte**

Bezieht man performanceorientierte Daten in die NLG-Software ein, kann sich der Prozess der Textautomatisierung selbst trainieren und verbessern.

Einsatzgebiete von automatisierter Texterstellung im Medienbereich

Automatisch generierte Texte werden im **Medienbereich** derzeit vor allem in der **Finanzberichterstattung, bei regionalen/lokalen Sportevents und bei Wahlberichterstattung** sowie im Bereich **Service-News** eingesetzt.

▶ **Berichterstattung**

- ▶▶ Sport: z.B. Fußballergebnisse
- ▶▶ Sicherheit: z.B.: Verbrechen-Reports
- ▶▶ Finanzen & Wirtschaft: z.B. Börse-Reports, Geschäftsberichte
- ▶▶ Society: z.B. Prominews
- ▶▶ Politik: z.B. Wahlergebnisse, Wahlprognosen

▶ **Service**

- ▶▶ Umwelt: z.B. Wetterberichte, Feinstaub-Updates
- ▶▶ Verkehr: z.B. Staumeldungen, Ankündigung von Straßensperren
- ▶▶ Termine: z.B. Eventankündigung
- ▶▶ Branchenbuch: z.B. Firmendarstellungen

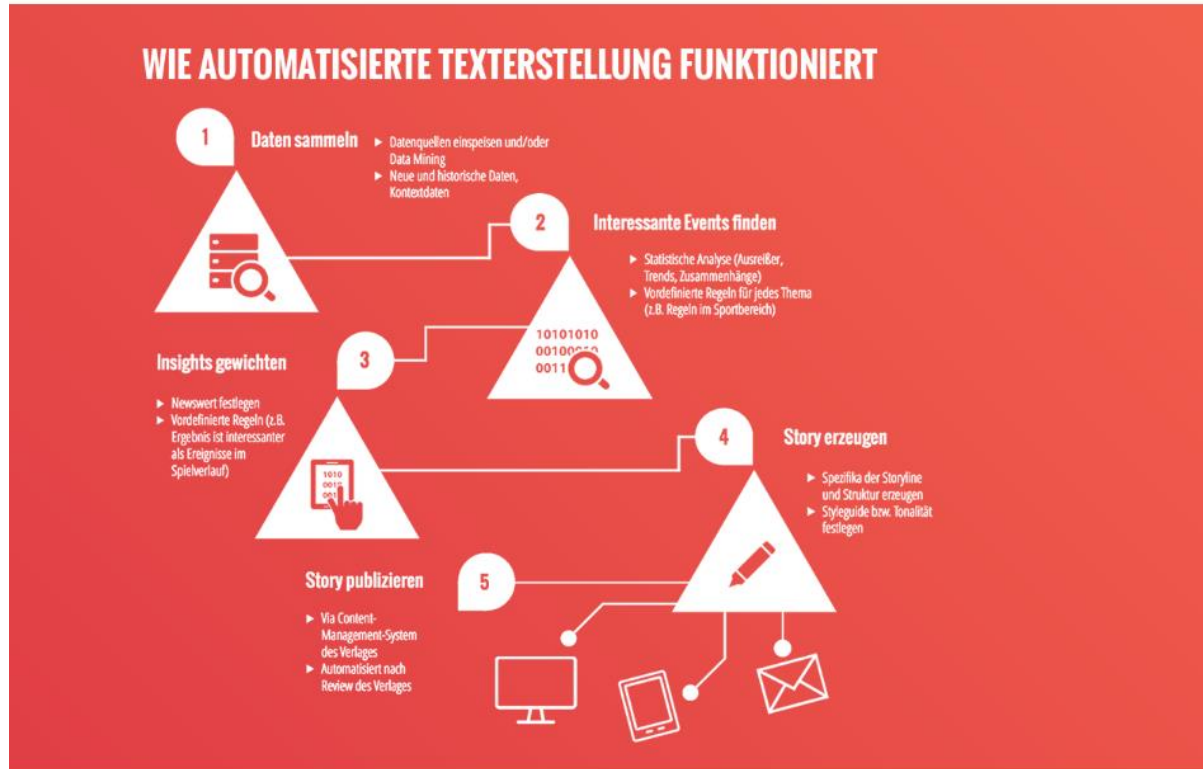
▶ **E-Commerce**

- ▶▶ Textierung von Angeboten in verlagseigenen Online-Shops
- ▶▶ Verknüpfung von Berichterstattung und kommerziellen Datenbanken z.B. aus dem Segment Essen&Trinken

▶ **Zukünftige Einsatzgebiete**

- ▶▶ Berichterstattung: Musik- und Buchvorstellungen
- ▶▶ Serviceinhalte: TV-Programm, Horoskope, Ausgehtipps
- ▶▶ Classifieds: Textierung von Inseraten
- ▶▶ Verlagsmarketing: Erstellung von individuellen Briefen für Abonnenten
- ▶▶ Neuartige Zusatzservices: Applikationen für Haus & Garten z.B. User geben Profil ihres Gartens ein (m2, Lage etc.) - Vorschläge zu Aktivitäten und Gestaltung rund um das Gartenjahr

Automatisierte Texterstellung: Der Prozess



Automatisierte Texterstellung: Knackpunkte und Limitationen

▶ **Datenverfügbarkeit und -qualität**

Für den journalistischen Bereich werden hochqualitative, verlässliche Daten in strukturierter und maschinen-lesbarer Form benötigt. Überall dort wo die Datenqualität schlecht oder die Verfügbarkeit nicht gewährleistet ist z.B. User Generated Information, niedriger Service Level bei Open Data, können die Ergebnisse verfälscht werden.

▶ **Validierung**

Algorithmen können zwar Korrelationen herstellen, aber nicht Kausalitäten erklären. Daher können nur Fakten berichtet werden, aber nicht warum bestimmte Dinge passieren. Die klassischen Tugenden des Journalismus: Einordnung und Kontextualisierung sind weiterhin dem Menschen vorbehalten.

▶ **Qualitätssicherung**

ist nur schwer zu bewerkstelligen. Es kann nicht mehr jeder Text gemonitort werden. Das kommt vor allem dort zum Tragen, wo die Technologie nicht als Tool im Newsroom genutzt wird, sondern als Ersatz für den Menschen, etwa durch die massenhafte Generierung zusätzlicher Texte.

▶ **Narration**

NLG ist derzeit noch am Beginn, schön geschriebene längere Texte können also noch nicht erwartet werden. Hier wird sich aber in den nächsten Jahren viel verbessern.

Marktüberblick- IT-Dienstleister für automatisierte Texterstellung

Firma	Land	Gründung	Launch NLG Software	Sprachen (Beispiele)	Themengebiete	Journalistische Produkte	Kunden (Beispiele)
Automated Insights	USA	2007	2014 Wordsmith Plattform	ENG	Finanzen, Sport	<ul style="list-style-type: none"> Finanzen: Quartalsberichte NCAA College Sportberichte Wahlberichterstattung 	<ul style="list-style-type: none"> AP (Sport, Finanzen) PA (Wahlen, Sport) Yahoo (Sport)
Narrative Science	USA	2010	2011 Quill Plattform	ENG	Finanzen, Sport	<ul style="list-style-type: none"> Finanz- und Börseberichte Sportstatistiken (Basketball, American Football, Softball, Baseball) 	<ul style="list-style-type: none"> Forbes (Finanzen) Game Changer (Sport) ~ 10 weitere Kunden (nicht öffentlich)
AX Semantics	GER	2001	2009 AX Semantics	ENG, DEU, FRA, SPA, NLD, DAN, SWE, NOR, ITA, POR	Sport, Entertainment, Finanzen, Wetter	<ul style="list-style-type: none"> Fußballberichte Finanzberichte Promi-Fußball News 	<ul style="list-style-type: none"> Nordwestzeitung Sportnachrichten.at SID ~ 5 weitere Kunden (nicht öffentlich)
Text-On	GER	2013	2014 Text-On	DEU	Finanzen	<ul style="list-style-type: none"> Finanzberichte Zusammenfassung von Handelsregisterauszügen 	<ul style="list-style-type: none"> Finanzen100.de 3 weitere Pilotprojekte (nicht öffentlich)
2txt NLG	GER	2013	2013 2txt	DEU	Finanzen, Sport	<ul style="list-style-type: none"> Fußballberichte Finanzberichte 	Verhandlungen
Retresco	GER	2008	2013 rtr Text Engine	DEU, ENG	Sport, Wetter, Finanzen	<ul style="list-style-type: none"> Fußball (Bundesliga bis Kreisklasse) Wetterberichte 	<ul style="list-style-type: none"> FussiFreunde (Radio Hamburg) Neue Osnabrücker Zeitung (Fußball) Weser-Kurier (Fußball)
Textomatic	GER	2015	2015 Textomatic	DEU, ENG, SPA, NLD, FRA, ITA, POL, GRC, SWE, POR	Sport, Finanzen, Reise, Wetter	<ul style="list-style-type: none"> Fußballberichte Börseberichte Reisetipps Wetterberichte 	<ul style="list-style-type: none"> HandelsBlatt (Finanzen) Berliner Morgenpost (Finanzen) Focus (Wetter) Google DNI Projekt: http://www.3dna.news/
Syllabs	FR	2006	2012 Data2Content	ENG, FRA, SPA	Politik	Wahlberichterstattung	<ul style="list-style-type: none"> Le Monde (Wahlen) L'Express (Wahlen) Le Parisien (Wahlen)
Labsense	FR	2011	2013 Scribt Engine	FR, ENG, DEU	Wirtschaft	Projekt Lokalberichterstattung	Verhandlungen
Arria	GB	2011	2012 Arria NLG Engine	ENG	Wetter	Wetterberichte	Meteo Group
Tencent	CHN	1998	2015 Dreamwriter Plattform	CHN	Finanzen	Wirtschaftsberichte	keine Infos
Identrics	BGR	2015	2015 Identrics	DEU, ENG, FRA, ITA, SPA, POR, Osteuropäische und Nordische Sprachen	Finanzen, Wirtschaft	Wirtschaftsberichte	SEE News



Checkliste: Welche Fragen müssen sich Medienunternehmen stellen?

- ▶ Welche Geschäftsbereiche eignen sich, um mit automatischer Texterstellung Mehrwert für die Kunden/Leser zu stiften?
- ▶ Welche Prozesse in der Inhalte-Erstellung können automatisiert werden? Wie sieht eine Zusammenarbeit von Mensch + Maschine aus?
- ▶ Wer sind meine Partner für die Produktion?
- ▶ Wie kann ich passende Datenquellen identifizieren und anzapfen?
- ▶ In welchem Bereich lohnt es sich selbst Datenbanken aufzubauen, die lizenziert werden können?
- ▶ Wie sieht die Refinanzierung aus?

Rückfragen zum aktuellen Briefing:

verena.krawarik@apa.at

Weitere Trendbriefings aus dem APA-medialab:

<https://medialab.apa.at/>

<https://www.facebook.com/apamedialab/>

