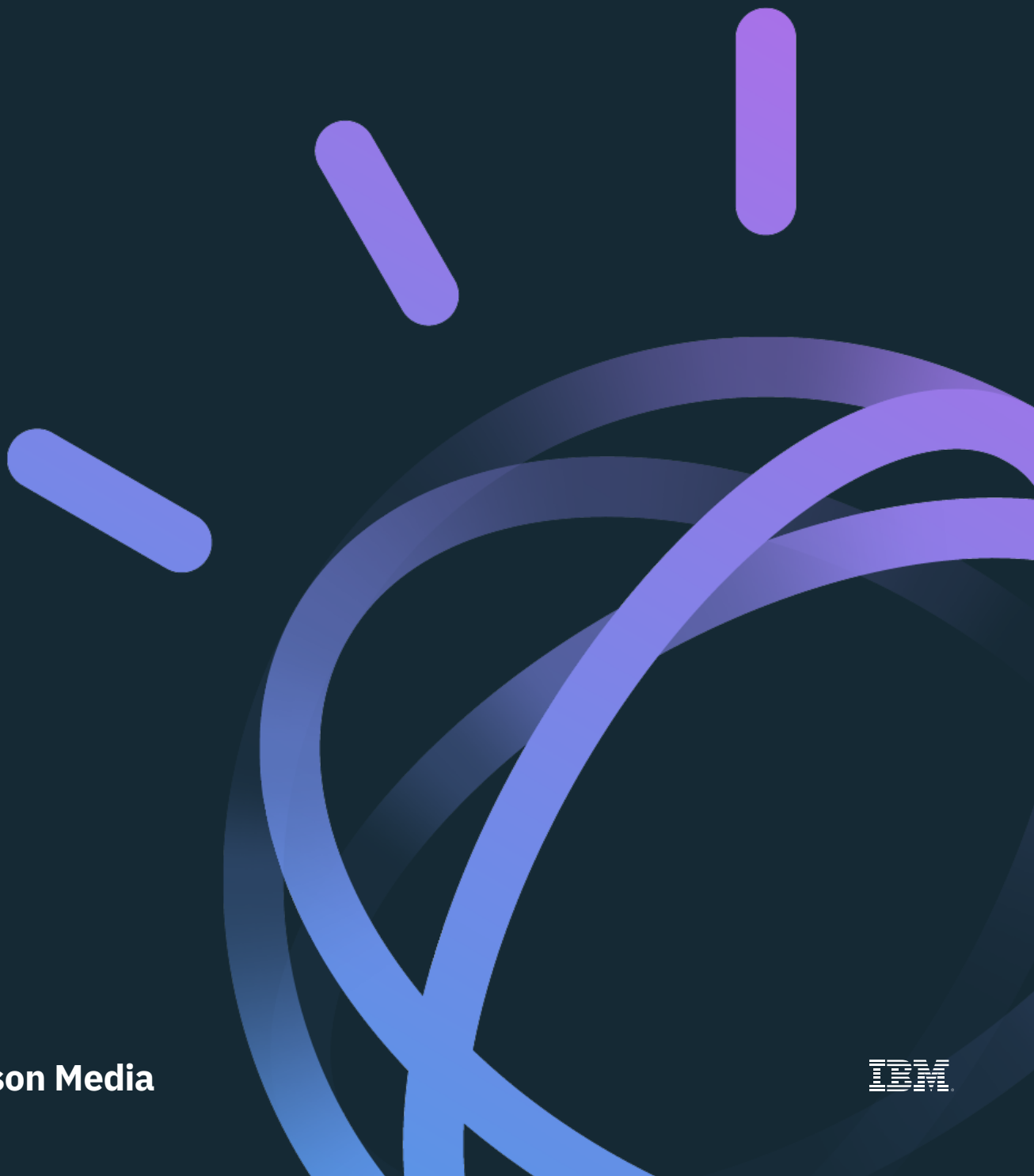
Whitepaper

Watson bei der Arbeit

Untertitelung wird kognitiv: ein neuer Ansatz für eine alte Herausforderung



IBM **Watson Media**

IBM

Kognitive Systeme versprechen einen größeren Kontext und eine verbesserte Genauigkeit zu wettbewerbsfähigen Preisen

Problem:
Die Untertitelung von Videos bleibt eine Herausforderung. Neue Vorschriften können das noch verschärfen.

Kontext:
Selbst mithilfe von automatisierten Systemen und Spezialisten von Drittanbietern ist die Untertitelung von Live- und Postproduktionen eine nach wie vor unvollkommene Kunst, die hartnäckig auf ihren Durchbruch wartet. Neue Anforderungen an die Untertitelung bestimmter Online-Videoinhalte werden noch größere Anforderungen an eine wirtschaftliche und präzise Untertitelung stellen.

Lösung:
Kognitive Systeme haben das Potenzial, neue Fähigkeiten bei der Untertitelung einzubringen, insbesondere im Bereich schnelle Erfassung und Anwendung von kontextuellem, menschenähnlichem Verständnis, um Fehler zu reduzieren und die Verständlichkeit von Untertiteln zu verbessern.

„Kognitive Systeme können dort erfolgreich sein, wo bisherige Plattformen für eine Untertitel-Automatisierung versagt haben."

Der erste allgemein anerkannte Fall von Untertitelung mit „Closed-Caption" war die 1972 von WGBH-TV ausgestrahlte Kult-Kochsendung „The French Chef" mit der verstorbenen Julia Child. Seitdem hat es eine Reihe von Initiativen gegeben, die die mühsame Umsetzung von gesprochenen Worten und Tönen in textuelle Beschreibungen und Untertitel automatisieren oder zumindest erleichtern sollen.

Auf externe Untertitelungsdienste, die Rundfunkanstalten und Netzwerke entlasteten, folgte die Einführung einer Software, die darauf abzielte, Töne und Sprache zu erkennen. Für dringendere Anforderungen kam 1982 die Echtzeit-Untertitelung zum Einsatz, als das National Captioning Institute einen versprechenden Dienst einweihte, der innerhalb von vier bis fünf Sekunden nach der Ausstrahlung Textinterpretationen von Worten und Tönen zurückgeben sollte. Seine geheime Zutat: Horden von flinken Berichterstattern, die Kopfhörer tragen und innerhalb von vier bis fünf Sekunden nach der Ausstrahlung fast-Live-Text produzieren.

Alle diese Teilnehmer haben den Stand der Technik bei der Untertitelung erhöht. Es war aber nichts ideal. Die „Live"-Untertitelung hat zahlreiche Faux-Pas hervorgebracht, während Schreibkräfte darum kämpften, unbekannte Wörter, Namen und Ausdrücke zu verstehen und in eine sofortige Ausgabe umzuwandeln. Softwarelösungen waren in der Vergangenheit ebenfalls unvollkommen, und das oft aus den gleichen Gründen. Rechtschreibfehler, „unhörbare" Begriffe und vor allem Schwierigkeiten bei der Interpretation des Kontextes gesprochener Worte haben die Geduld von Kunden aus der Fernsehindustrie auf die Probe gestellt, die gezwungen sind, fehlerhafte Ergebnisse zu überprüfen und zu korrigieren, bevor sie Untertitel in ihre fertigen Videodateien integrieren.

Daher ist es verständlich, dass sich unter Fachleuten der Videobranche oft ein gewisses ernüchterndes Gefühl „Kennt man schon" breit macht, denen man viele Lösungen versprochen hat, nur um das eigentliche Handwerk auszuführen, das es immer noch ist: ein manuell aufwendiger und beharrlich unvollkommener Prozess.

Durchbruch kognitiver Systeme

Heute hingegen kann die Untertitel-Szene von einer neuen Möglichkeit profitieren, die das Potenzial für einen echten Durchbruch hat. Kognitive Systeme kombinieren die bereits vorhandene Videoverarbeitungsintelligenz mit etwas Neuem – der Fähigkeit, den umgebenden Kontext von Videoinhalten zu analysieren, zu verstehen und zu „lernen", ähnlich, wie es Menschen tun. So wird das Wort „Fehler" im Zusammenhang eines Tennisspiels anders behandelt, als es bei einer Seifenoper-Episode der Fall wäre. Und die sich ergebende Interpretation und Darstellung von Wörtern und Beschreibungen, die diesem Wort vorausgehen und ihm folgen, werden präziser und mit einer wesentlich verbesserten kontextuellen Darstellung wiedergegeben. Kognitive Systeme haben die Chance, dort erfolgreich zu sein, wo bisherige Plattformen für die Untertitel-Automatisierung versagt haben. Denn sie erzeugen einen Output, der die Absicht, den Zweck und die wortgetreue Zusammenstellung von Wörtern und Tönen, die an Videoinhalte gebunden sind, genauer verfolgt. Da diese Systeme untersuchen und interpretieren können, arbeiten sie fast genauso gut wie menschliche Transkriptionsspezialisten. Nur, sie sind schneller.

„Durch das Selbstlernen bei jeder Korrektur wird die Erkennungsgenauigkeit bei jedem Einsatz verbessert.“

Die Schlüsseltechnologien, die Watson für die Untertitelung einsetzt, sind:

- **Anpassung des Sprachmodells** erstellt domänenspezifische Sprachmodelle, um die Erkennungsgenauigkeit zu erhöhen
- **Benutzerdefiniertes Korpus** erweitert den Wortschatz mit Wörtern im Kontext
- **Benutzerdefinierte Wörter** erweitern den Wortschatz mit Wörtern und deren phonetischer Form
- **Benutzerdefinierte Akustikmodelle** verbessern die Erkennungsgenauigkeit für Videos mit spezifischen Audiobedingungen (Hintergrundgeräusche, spezielle Akzente usw.)

Untertitel powered by Watson

Watson von IBM nutzt automatisierte Spracherkennungsfunktionen, um gesprochene und auditive Elemente von Videoinhalten zu erfassen. Das Programm wendet dann eine Reihe von kognitiven Funktionen an, um die interpretierten Daten zu bewerten und darauf zu reagieren. Darüber hinaus ermöglicht Watson maßgeschneiderte Untertitellösungen dank Funktionen wie Korpus, Vokabular und benutzerdefinierte Audiomodelle, um die Genauigkeit von Untertitelskripten beim ersten Durchlauf weiter zu verbessern.

Watson generiert automatisch Untertitel für aufgenommene Videos mithilfe der „Watson Speech to Text“-API. Die Caption-Editor-Funktion der API wurde entwickelt, um die automatisch generierten Untertitel zu überprüfen und zu korrigieren. Die Editor-Oberfläche ist sowohl für Experten als auch für Laien konzipiert und auf maximale Effizienz optimiert. Durch das Selbstlernen bei jeder Korrektur wird die Erkennungsgenauigkeit bei jedem Einsatz verbessert. Namen und Eigennamen werden automatisch aus überprüften Untertiteln extrahiert und in Glossare übernommen, um sicherzustellen, dass sie in nachfolgenden Einsätzen richtig erkannt und geschrieben werden. Die Untertitel werden mithilfe eines intelligenten Layout-Algorithmus erzeugt. Mit diesem Algorithmus segmentiert Watson automatisch Untertitel an natürlichen Bruchstellen, was zu besser lesbaren Untertiteln führt.

Die Watson und US Open Demo zeigt, wie Untertitel automatisch mithilfe der „Watson Speech to Text“-API generiert werden.

Watson fügt Untertitel in die US Open ein (Demo-Umgebung)

Die Geldfrage

Natürlich müssen sich kognitive Systeme auch wirtschaftlich bewähren, indem sie nutzbare Ergebnisse zu gleichen oder niedrigeren Kosten produzieren als die vorherrschende Industrienorm. Am oberen Ende der Skala können die Ausgaben für die Untertitelung von Live-Inhalten zwischen 3 und 5 US-Dollar pro Minute liegen, einschließlich automatisch generierter Skripte und menschlicher Bearbeitung; der untere Bereich für Inhalte, die nicht live und Video-on-Demand sind, liegt bei etwa 1 bis 2 US-Dollar pro Minute. Dies sind die Grenzen, die kognitive Systeme erfüllen oder übertreffen müssen, um attraktiv zu sein.

Die gute Nachricht an dieser Stelle ist, dass kognitive Systeme Teilnehmer an einem wachsenden Markt für automatische Inhaltserkennung (ACR) sind, mit positiven Auswirkungen bei großen Volumen. Das globale Marktforschungsunternehmen MarketsandMarkets prognostiziert eine durchschnittliche jährliche Wachstumsrate von 27,2 % bis 2021 für die ACR-Technologie im Medien- und Unterhaltungssektor, die von einem Zusammenströmen von Anwendungen im Bereich Audio-Fingerprinting, Markierung, Musikerkennung und Musikentdeckung angetrieben wird. Im Wesentlichen wird die steigende Nachfrage, personalisierte Medieninhalte besser zu finden und zu entdecken, dazu führen, dass die Kosten für ACR-Technologien auf mehr Sektoren aufgeteilt werden, was zu Kostenverbesserungen für Untertitelungsnutzer (neben vielen anderen Anwendern) führen kann. Auch die Erweiterung des Markts für Video-Untertitelung trägt zu einem größeren Umfang bei. Die im Juli 2017 in Kraft getretenen neuen US-Untertitelungsregeln schreiben vor, dass Online-Videoinhalte, die ursprünglich „im Fernsehen gezeigt“ wurden, mit Untertiteln versehen werden müssen. Diese Anforderung hat wiederum das Potenzial, die Gesamtmarktgröße für Untertitel zu vergrößern, wodurch ein größerer Investitionspool zur Verfügung gestellt werden kann, um die Kosten zu senken.

Watson nutzt automatisierte Spracherkennungsfunktionen, um gesprochene und auditive Elemente von Videoinhalten zu erfassen.

Genauigkeit, Genauigkeit, Genauigkeit

Natürlich ist der andere wichtige Gesichtspunkt für die Videoindustrie die Einhaltung von Gesetzesvorschriften. In den USA verlangt die FCC, dass Untertitel Folgendes erfüllen müssen:

- **Genauigkeit:** Untertitel müssen mit den gesprochenen Worten des Dialogs übereinstimmen und Hintergrundgeräusche und andere Töne so weit wie möglich wiedergeben.
- **Synchronisation:** Untertitel müssen so weit wie möglich mit den entsprechenden gesprochenen Worten und Tönen übereinstimmen und in einer für den Betrachter lesbaren Geschwindigkeit auf dem Bildschirm dargestellt werden.
- **Vollständigkeit:** Untertitel müssen vom Anfang bis zum Ende des Programms im größtmöglichen Umfang angezeigt werden.
- **Richtige Platzierung:** Untertitel sollten andere wichtige visuelle Inhalte auf dem Bildschirm nicht blockieren, sich nicht überlappen oder über den Rand des Videobildschirms hinaus reichen.

Kognitive Systeme können zu den meisten dieser Grundvoraussetzungen beitragen, da sie die Audioerkennung mit einem breiteren Kontextverständnis verknüpfen können, weshalb die den Redakteuren zur abschließenden Überprüfung vorgelegten Transkripte genauer und marktreifer als Vorgängertechnologien sind.

Implementierungsszenarien

Teilnehmer aus der Videoindustrie, die den Beitrag kognitiver Systeme bei der zukünftigen Untertitelung verstehen wollen, können erste Implementierungen testen und sich so einen ersten Eindruck verschaffen. Für die Rechtfertigung erster Versuche können folgende Überlegungen zählen:

- Die Gesamtkosten sind ähnlich oder niedriger als bei verfügbaren Lösungen
- Inhouse-Lösungen werden bevorzugt
- Die Bearbeitungszeit ist entscheidend
- Inhalte, für die keine Vorschriften gelten

In der Fernsehindustrie befinden sich kognitive Systeme in einem frühen Stadium. Es besteht jedoch ein großes Interesse an einer Lösung, die die Einschränkungen und Grenzen von alten Ansätzen überwindet. Die leistungsstarke Kombination aus automatischer Inhaltserkennung und kognitiven/lernenden Fähigkeiten wird neue Möglichkeiten in die langjährige Praxis der Fernsehindustrie einbringen. Am Ende können kognitive Systeme genau das hervorbringen, was die Untertitelung erreichen wollte, seit Julia Child der Welt gezeigt hat, wie man wie ein französischer Meisterkoch kocht.

IBM Cloud Video
550 Kearny Street, Suite 600
San Francisco, CA 94108.

Hergestellt in den Vereinigten
Staaten von Amerika
Dezember 2017

Die in diesem Dokument enthaltenen Informationen sind auf dem Stand des Datums der Veröffentlichung und können jederzeit von IBM geändert werden. Nicht alle Angebote sind in allen Ländern verfügbar, in denen IBM vertreten ist.

Die Informationen in diesem Dokument werden im vorliegenden Zustand ohne jegliche ausdrückliche oder implizierte Garantie und ohne Garantien für Marktgängigkeit und Eignung für einen bestimmten Zweck bzw. ohne Garantien oder Bedingungen der Nichtverletzung bereitgestellt.

Die Garantie der Produkte von IBM fällt unter die Geschäftsbedingungen der Verträge ihrer Bereitstellung.

Erklärung zu bewährten Praktiken im Sicherheitsbereich: IT-Systemsicherheit impliziert das Schützen von Systemen und Informationen durch Prävention, Ermittlung und Reaktion auf unerlaubten Zugriff innerhalb und außerhalb des Unternehmens. Unerlaubter Zugriff kann dazu führen, dass Informationen verändert, zerstört, oder missbraucht werden. Er kann zu Schaden oder Missbrauch Ihrer Systeme führen, zu dem auch Angriffe gegen andere gehören können. Kein IT-System oder Produkt sollte als vollständig sicher angesehen werden und kein Produkt bzw. keine Sicherheitsmaßnahme kann bei der Vermeidung von unerlaubtem Zugriff vollständig wirksam sein. Die Systeme und Produkte von IBM wurden als Teil einer umfassenden Sicherheitsmethode konzipiert, zu der unbedingt weitere Verfahren gehören und die andere Systeme, Produkte und Service erfordern kann, um optimal wirksam zu sein.

IBM bietet keine Garantie, dass Systeme und Produkte gegen bösartiges oder illegales Verhalten Dritter immun sind.