Schema.org im Tourismus

von Elias Kärle und Dieter Fensel, Semantic Technology Institute, Universität Innsbruck, Österreich {elias.kaerle, dieter.fensel}@ sti2.at über den Weg mit semantischen Technologien zu mehr Sichtbarkeit und Umsatz im World Wide Web zu gelangen.



Abstract

Mit der Verbreitung von schema.org gelang es erstmals das "Semantic Web" einer großen Masse von Anwendern verständlich zu machen. Dieser Vorstoß kann auch im Tourismus von großem Nutzen sein. In einer Zeit, in der die Direktvermarktung zu Gunsten großer Buchungsplattformen auf der Strecke bleibt und Hotelwebseiten einzelner Anbieter kaum noch in Suchmaschinen

gefunden werden können, bringen semantische Technologien einen entscheidenden Wettbewerbsvorteil und können, richtig verwendet, zu mehr Sichtbarkeit und Umsatz führen. Dieser Artikel erklärt, wozu schema.org in der Lage ist, wie diese Technologie dem Tourismus helfen kann und wie das Semantic Technology Institute in Innsbruck an der Buchung der Zukunft arbeitet.

1. Einleitung

OTAs dominieren den Buchungsmarkt: Die Erfindung und Verbreitung des World Wide Web (WWW) in den späten 1990er Jahren, eröffnete Beherbergungsbetrieben erstmals die Möglichkeit, ihr Unternehmen und ihre Angebote digital einer breiten Masse vorzustellen. Täglich entstanden neue Webagenturen und Webseiten und dem neuen Vertriebsweg wurde eine große Zukunft prognostiziert. Allerdings blieben Beherbergungsbetriebe nicht die einzigen, die das Potential von Webseiten erkannten und so entstanden schnell große Buchungsplattformen, sogenannte Online Travel Agencies (OTAs), die heute den online Buchungsmarkt bei Beherbergungsbetrieben dominieren. Einerseits geben Buchungsplattformen wie booking. com, Expedia oder Metasuchmaschinen wie Trivago einen guten Überblick über verfügbare Zimmer und Zimmerpreise in einer Region und bringen somit auch kleineren Beherbergungsbetrieben, die sich kein aufwändiges Marketing leisten können, Sichtbarkeit und Gäste, die sie sonst nur schwer erreichen würden. Andererseits bringen genau jene großen Buchungsplattformen mit diversen Knebelverträgen, wie der "Bestpreisklausel", die Betriebe bei der Preisgestaltung und beim Wettbewerb untereinander in Bedrängnis, verlangen sehr hohe Kommissionen und verringern die Direktbuchungen über die Webseite eines Unternehmens enorm (Schegg & Fux, 2012)..

Das unsichtbare Hotel: Die Buchung des Kunden von einer der großen Buchungsplattformen auf die eigene Webseite zu verlagern, ist aus gegenwärtiger Sicht kaum oder nur sehr schwer möglich. Bei der Suche nach "Hotel" oder "Unterkunft" und einem Orts- oder Regionsnamen ist die erste Seite jeder Suchmaschine dominiert von Angeboten der Big Player wie booking.com, Trivago, Expedia oder anderen. Die eigene Webseite des Betriebes oder einer Region geht dabei völlig unter und wird schlimmstenfalls sogar erst auf der zweiten Seite der Suchergebnisse gereiht, was dazu führt, dass ein großer Teil der Benutzer der Suchmaschine diese Seite bei seiner Suche gar nicht zu Gesicht bekommt. Die

Suchmaschinenanbieter selbst sehen dabei nicht tatenlos zu und haben schon sehr früh das Potential im Tourismus erkannt. Abgesehen von dem "alten" Geschäftsmodell der bezahlten Werbung finden sich nun immer häufiger bei der Suche nach Beherbergungsbetrieben Bereiche auf der ersten Seite, wo direkte Buchungsmöglichkeiten über die jeweilige Suchmaschine angeboten werden.

Das "headless web": Es ist somit ein Trend im Web erkennbar, der sich weit über den Tourismus hinaus erstreckt. Man spricht von einer neuen Schicht, die über dem ursprünglichen World Wide Web entsteht, in welcher zum Beispiel Waren nicht mehr von einzelnen Shops verkauft werden, sondern von großen Plattformen wie Alibaba oder Amazon. Zimmer werden nicht mehr von Hotels oder Tourismusverbänden verkauft, sondern von großen Buchungsplattformen und Informationen werden nicht mehr von den jeweiligen Webseiten geliefert, sondern von den Suchmaschinen selber, welche die Informationen aus den Webseiten extrahieren.

Wir, am Semantic Technology Institute, arbeiten mit Hochdruck an Lösungen, um Betrieben zu helfen in diesem "headless web" Fuß zu fassen bzw. Oberwasser zu halten und Nächtigungen und Angebote über die eigene Webseite zu vertreiben. Deshalb stellen wir in diesem Artikel schema.org, eine Möglichkeit Daten auf Webseiten strukturiert und Maschinen les- und interpretierbar zu machen, und einen Vier-Punkte-Plan, der über diese Technologie aufklärt, zu ihrer Nutzbarkeit beiträgt und Umsätze durch deren Verwendung erzeugt, vor.

2. Schema.org

Was ist schema.org? Im Jahre 2011 entschlossen sich die vier "Big Player" am Suchmaschinenmarkt, Bing, Google, Yahoo! und Yandex, eine Union zu bilden, um Inhalte im Web besser zu strukturieren und somit für Suchmaschinen lesbar und verständlich zu machen (Abbildung 1). Die Initiative, die daraus entstanden ist, nannten sie schema.org. Schema.org definiert eine Sammlung von Termen um "Dinge" auf Webseiten zu beschreiben. So kann beispielsweise eine Hotelwebseite mit dem Term sche-

ma.org/Hotel "markiert" oder "annotiert" werden und zusätzlich besteht die Möglichkeit, eine ganze Reihe von Attributen anzugeben, die dieses Hotel beschreiben, wie zum Beispiel der Name, die Adresse, Webseite und vieles mehr. Da diese sogenannten Annotationen für den Benutzer der Webseite nicht sichtbar sind, spricht man hier auch oft von "Metadaten" die in den Quellcode integriert sind. Ein Beispiel, wie schema.org in eine Webseite integriert werden kann, finden Sie in Listing 1.



Abbildung: Die "Big 4" Erfinder von schema.org: Bing, Google, Yahoo! und Yandex.

Warum schema.org? Wie eingangs schon erwähnt verliert das Web wie wir es derzeit kennen, als eine Sammlung von Webseiten, bei denen sehr viel Wert auf die Optik und die Benutzerfreundlichkeit gelegt wird, immer mehr an Bedeutung. Vor allem bei Beherbergungsbetrieben rückt der Aspekt der Datenqualität immer mehr in den Mittelpunkt. Ausschlaggebend dafür sind zum Beispiel Suchmaschinen, die immer mehr zu Antwortmaschinen werden und keine Verlinkungen zu Webseiten mehr auflisten, sondern nur noch direkte Antworten geben, extrahiert aus strukturierten Daten (siehe Abbildung 2). Ein weiteres Beispiel ist der Vormarsch der "Intelligent Personal Assistent"², wie Apple Siri, Google Now, Microsoft Cortana oder andere. Auch Chatbots bekommen bei der Suche und Buchung von touristischen Angeboten immer mehr Relevanz und auch diese bekommen ihre Informationen aus strukturierten Daten im Web.

Weitere Gründe, die für eine Verwendung von schema.org auf Webseiten sprechen sind zum Beispiel Semantic SEO. Dies ist eine weit verbreitete und

¹ Der Begriff "headless web" wurde hier bewusst nicht übersetzt. Einen passenden deutschen Begriff dafür zu finden ist schwierig, da schon im englischen "headless web" unglücklich gewählt wurde. Es fehlt dem Web nämlich nicht der Kopf, sondern eigentlich die grafische Anzeige, der Monitor, siehe dazu: https://paul.kinlan.me/the-headless-web/ (Zugriff: 21.9.2016)

² https://en.wikipedia.org/wiki/Intelligent_personal_assistant

vielseitig verwendbare Technik, um Webseiten für Suchmaschinen zu optimieren. Da die Suchmaschine den Inhalt der Webseite nicht nur syntaktisch versteht, sondern auch semantisch interpretieren kann, steigt zwar die Sichtbarkeit nicht zwangsläufig, jedoch sinkt die Absprungrate auf Webseiten merklich. Google belohnt die Verwendung von schema. org außerdem noch mit Rich Snippets und Rich Cards. Dies sind ansprechende grafische Aufbereitungen der Produkte und Services einer Webseite, die direkt im Suchmaschinen Frontend eingebunden werden und so noch besser gefunden werden können. Bei Hotels und Beherbergungsbetrieben in den Vereinigten Staaten ist schema.org schon sehr weit verbreitet, in Europa besteht hier noch Aufholbedarf, aber es ist eine starke Tendenz erkennbar, die erahnen lässt, dass schema.org in Zukunft nicht mehr wegzudenken sein wird (Kärle, Fensel, Toma, & Fensel, 2016).

Wie verwendet man schema.org? Um Annotationen jeglicher Art in Webseiten zu integrieren gibt es grundsätzlich drei verschiedene Möglichkeiten: RDFa, Microdata und JSON-LD. Schema.org emp-

fiehlt die Verwendung von JSON-LD, deshalb werden die anderen beiden Formate auf dieser Stelle vernachlässigt. Das JSON-LD Format ist, wie der Name schon erahnen lässt, stark angelehnt an JSON und wird durch zusätzliche Terme für die Verwendung mit Linked Data (-LD) erweitert. Beispiele für solche zusätzlichen Linked Data spezifischen Terme sind @contex – um das verwendete Vokabular auszuzeichnen, @type – der verwendete Objekttyp oder @id – die eindeutige Identifikation eines Objekts. Der JSON-LD Baustein bzw. das Snippet, welches alle relevanten Inhalte der jeweiligen Seite beschreibt, wird dann in einem <script>-Tag direkt in den HTML Code der Webseite integriert. Ein Beispiel dazu in Listing 1.

3. Strategie

Am Semantic Technology Institute in Innsbruck verfolgen wir deshalb schon seit längerem eine aus vier Punkten bestehende Strategie, um semantische Technologien im Tourismus zu etablieren, nutzbarer zu machen und Anwendungen, die zu Umsätzen führen, zu erstellen.

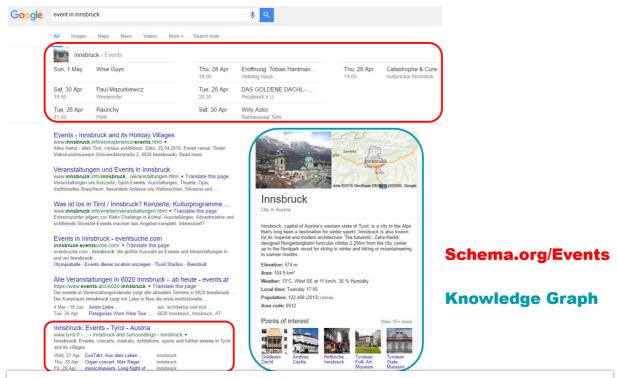


Abbildung: Informationen zu "Events in Innsbruck" werden direkt in Google angezeigt. Die Information dafür kommt aus schema.org annotierten Webseiten und dem Google Knowledge Graph.

Bewusstsein: Um Web-Agenturen, Tourismusverbänden, Betreibern von Beherbergungswebseiten und anderen Anbietern touristischer Services ein Verständnis zu geben, wie wichtig die Verwendung semantischer Technologien wie schema.org schon ist und welchen großen Stellenwert diese Technologien in der Zukunft haben werden, haben wir uns mit der österreichischen Hotelvereinigung (ÖHV), dem Bundesministerium für Wissenschaft Forschung und Wirtschaft (BMWFW) und der Technischen Universität Wien (Electronic Commerce Group, Institut für Softwaretechnik und Interaktive Systeme) zusammengetan, um gemeinsam einen Leitfaden zu erarbeiten. Darin widmet sich ein großes Kapitel der Wichtigkeit und der richtigen Verwendung von schema.org. Dieser Leitfaden wird nach Fertigstellung vom OHV und dem BMWFW an Web- und Werbeagenturen sowie an interessierte Hotelbetriebe verteilt und soll als eine Orientierungshilfe bei der Erstellung einer touristischen Webseite dienen. Agenturen erfahren darin, worauf es bei einer stateof-the-art Webseite ankommt und Beherbergungsbetriebe können anhand der Informationen konkrete Wünsche an Agenturen richten und diese dann auch überprüfen.

Hilfestellung: Um nicht nur darauf hinzuweisen was wichtig und nötig ist, sondern um auch aktive Hilfestellung zu leisten, arbeiten wir derzeit an zwei verschiedenen Softwarelösungen, die die Verwendung von schema.org erleichtern sollen. Der "schema.org Generator" ist eine Webanwendung, mit deren Hilfe Benutzer in der Lage sind, schema.org Snippets (Code Stücke) zu erzeugen. Der Benutzer wählt dabei das zu annotierende Objekt (z.B.: Hotel, Event, Person ...) und wird dann von der Software durch den Eingabeprozess geleitet. Am Ende des Vorganges wird das fertige Code Stück angezeigt und kann vom Benutzer direkt in die eigene Webseite kopiert werden. Dadurch ist es auch in Programmierung oder JSON-LD unerfahrenen Personen möglich, valide schem.org Annotationen zu erzeugen. Das zweite Werkzeug, an dem wir arbeiten ist der "schema.org Validator". Diese Webanwendung bietet Hilfestellung bei der Überprüfung der Richtigkeit annotierter Webseiten oder Codeteile. Im Gegensatz zu schon existierenden Testwerkzeugen, wie dem Structured Data Testing Tool von Google, validiert diese Software nicht nur die Syntax der Annotationen, sondern prüft auch semantische Richtigkeit mit der Hilfe von Regeln die von Domainexperten vorab festgelegt werden. So kann zum Beispiel überprüft werden, ob die angegebene Telefonnummer mit dem angegebenen Land übereinstimmt, oder ob die angegebene Zimmergröße einer Mindestfläche entspricht. Die Verwendung ist sehr einfach, der Benutzer braucht lediglich die URL der zu prüfenden Seite oder das zu überprüfende Quellcode Segment in das dafür vorgesehene Eingabefeld kopieren und die Validierung auf Knopfdruck star-

```
2.
        "@context": "http://schema.org/",
3.
        "@type": "Hotel",
        "@id": "https://www.landgasthof-adler.at"
4.
        "name": "Landgasthof Adler",
5.
        "openingHours": "We-Mo 10:00-20:00",
6.
        "address": {
7.
            "@type": "PostalAddress",
8.
9.
            "addressCountry": "Austria",
            "addressLocality": "Hinterhornbach",
10.
            "addressRegion": "Tirol",
11.
            "postalCode": "6642",
12.
            "streetAddress": "Hinterhornbach 17"
13.
14.
15.
        "email": "office@landgasthof-adler.at",
        "telephone": "+435632318",
16
        "description": "Landgasthof Adler, Tiroler Wirtshaus, Hinterhornbach. Tiroler Spezi
17.
    alitäten und Zimmer im Herzen der Allgäuer Alpen.",
18.
        "url": "http://www.landgasthof-adler.at",
19. }
```

Listing 1: Ein einfaches JSON-LD Beispiel eines Hotels

ten. Somit kann zum einen überprüft werden, ob die eigenen Annotationen richtig sind und zum anderen bietet dieses Werkzeug eine Qualitätskontrolle über von Webagenturen erbrachte Leistungen.

Ausdrucksstärke: Das Ziel der Entwicklung von Schema.org war, eine möglichst ausdrucksstarke aber doch einfach zu benutzende Sammlung von Termen zu erschaffen, um eine möglichst breite Masse an "Dingen" im World Wide Web beschreiben zu können. Dieses ist in vielerlei Hinsicht sehr gut gelungen, jedoch war die Ausdrucksstärke von schema.org im touristischen Bereich eher bescheiden. Bis zur Version 3.0 war es möglich Stammdaten von Hotels wie Name, Beschreibung, Adresse und dergleichen zu beschreiben, jedoch gab es keine Möglichkeit angebotene Zimmer oder Zusatzfeatures wie eine Sauna oder einen Fitnessraum zu annotieren. Deshalb haben wir, in Zusammenarbeit mit Prof. Martin Hepp (E-Business und Web Science Research Group, Universität der Bundeswehr München, Deutschland), an einer Erweiterung des Vokabulars gearbeitet. Diese Erweiterung umfasst 12 neue Typen (Appartement, Suite, Konferenzraum, Bettdetails ...) und 10 neue Eigenschaften (Rauchen erlaubt, Belegung, Zimmergröße...). Nach eingehender Prüfung und mehreren Verbesserungsvorschlägen durch das "schem.org steering committee" wurde die Erweiterung im August 2016 in die Version 3.1 von schema.org aufgenommen und ist somit offiziell anerkannt und verwendbar (Kärle, Simsek, Hepp, & Fensel, 2016).

Umsatz: Das Ziel jedes Beherbergungsbetriebes ist es, möglichst viele Zimmer zu vermieten und dabei bestmögliche Preise zu erzielen. Direktbuchungen über die hoteleigene Webseite stagnieren jedoch, während Buchungen über OTAs weiter im Aufschwung sind (Schegg & Fux, 2012). Um die größtmögliche Reichweite bei kleinstmöglicher Kommissionszahlung zu erreichen und das Ungleichgewicht zwischen Direktvertrieb und OTAs umzuverteilen, arbeiten wir an einem Projekt, um Direktbuchungen voranzutreiben. ADB (Automated Direct Booking) basiert auf der Verfügbarkeit aller buchungsrelevanten Daten auf der Homepage

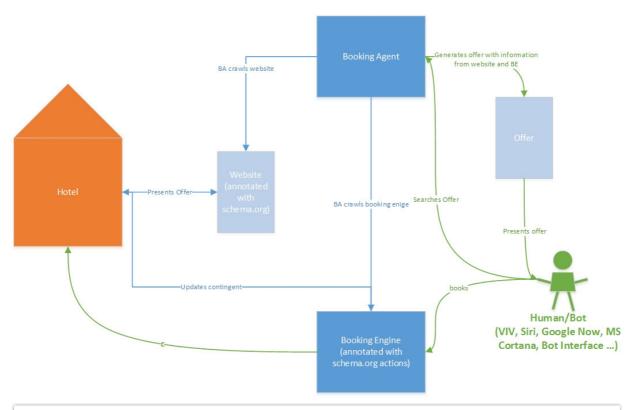


Abbildung 3: Schematische Darstellung von ADB.

eines Hotels, annotiert mit schema.org. Von dort werden die Daten gesammelt, an einer zentralen Stelle (das kann zum Beispiel ein Tourismusverband sein) gespeichert und aufbereitet. Diese zentrale Stelle (der Booking Agent) bietet dann eine Schnittstelle (API) an die von externen Diensten angesprochen werden kann. Solche externen Dienste können sowohl intelligente persönliche Assistenten sein (Cortana, Google Now, Siri) oder auch Chatbots (Facebook M, Slack Bot) und ähnliche externe Dienste. Ziel ist es, eine Buchung möglichst einfach oder gar automatisch zu erledigen, dass der Kunde beispielsweise auf seinem Smartphone nur noch wenige Eingaben tätigen muss und alles andere von der Software erledigt wird. Die Idee des Booking Agents unterstützt dabei die Direktvermarktung über die hoteleigene Webseite und erspart dem Betreiber eine Integration seiner Daten in eine Buchungssoftware, da alle Daten schon als "Open Date" auf der Webseite verfügbar sind, dank schema.org Annotation. Für eine schematische Darstellung von ADB siehe Abbildung 3.

4. Zusammenfassung und Ausblick

Unsere gegenwärtigen Arbeiten an der Verbreitung und Nutzbarmachung semantischer Technologien im Tourismus stehen erst am Anfang eines langen Weges, weg vom klassischen Web, hin zu neuen Technologien und Ideen der Präsentation und Vermarktung eines Unternehmens, abseits großer Buchungsplattformen. Hinsichtlich des oben genannten Vier-Punkte-Plans gibt es noch jede Menge Raum für Erweiterungen. Der Leitfaden muss ständig auf dem Stand der Technik gehalten werden. Das Vokabular von schema.org kann beliebig erweitert werden. Zum Beispiel ist für die Zukunft eine Erweiterung geplant, um auch Skigebiete, Lifte und Pisten zu annotieren. Die beschriebene Software kann um neue Features und Funktionen ergänzt werden und das Projekt "Automated Direct Booking", derzeit nur für Zimmerbuchungen nutzbar, kann auf verschiedenste Arten von touristischen Services erweitert werden.

Abschließend bleibt zu sagen, dass der Tourismus mit semantischen Technologien in eine spannende Zukunft blickt. Es ist jedoch wichtig, dass jeder Teil dieser Wertschöpfungskette die Zeichen der Zeit erkennt und sich diesen neuen Ideen nicht verschließt. Eine automatische Buchung auf Basis offen zugänglicher Daten funktioniert nur, wenn Hotelbetreiber dieses Daten zur Verfügung stellen.

Und dies wiederum ist nur möglich, wenn die passende Aufklärungsarbeit betrieben wird und wenn das nötige Vokabular und die nötigen Werkzeuge dafür zur Verfügung stehen. Wir vom Semantic Technology Institute in Innsbruck arbeiten mit Hochdruck an Lösungen und freuen uns immer über aufgeschlossene Tourismusverbandschefs oder Betreiber von Beherbergungsbetrieben, die sich bereit erklären, unsere Technologien einzusetzen und sich als Pilotprojekte zur Verfügung stellen. Dadurch kann diese Synergie aus Forschung und Tourismus erst entstehen und langfristige, nachhaltige Erfolge bringen.

Referenzen:

Kärle, E., Fensel, A., Toma, I., & Fensel, D. (2016). Why Are There More Hotels in Tyrol than in Austria? Analyzing Schema.org Usage in the Hotel Domain. In Information and Communication Technologies in Tourism 2016 (pp. 99–112). incollection, Springer.

Kärle, E., Simsek, U., Hepp, M., & Fensel, D. (2016). Extending the schema . org vocabulary for more expressive accommodation annotations. Not yet Published.

Schegg, R., & Fux, M. (2012). The Power of Internet Distribution Systems (IDS). Institute for Tourism of the University of. Retrieved from http://www.etourism-monitor.ch/sites/default/files/downloads/distribution_survey_d-a-ch_scheggfux_2012_eng.pdf

Danksagung: Die Autoren bedanken sich bei der Online Communication Working Group des STI-Innsbruck für die Zusammenarbeit und die vielen konstruktiven Diskussionen während der wöchentlichen Meetings und beim Organisationskomitee der Tourism Fast Forward Konferenz für die Einladung und die Möglichkeit, dieses Thema vorgestellt zu haben.