

Thomas Grübel

# WAS SEAS VON SEOS LERNEN KÖNNEN

Der Artikel gibt eine Reihe praktischer Beispiele und Impulse, wie SEO-Tools den Alltag eines SEA-Kampagnen-Managers erleichtern können. Insbesondere werden Crawling- und Scraping-Techniken vorgestellt, um Websites strukturell und inhaltlich zu erfassen. Die so gewonnenen Informationen können zum Kampagnen-Set-up, Ausbau, zum Check und zur Optimierung von Kampagnen verwendet werden.

DER AUTOR



Thomas Grübel – seit 2003 im Bereich Search tätig – ist Geschäftsführer von InBiz. Thomas berät Unternehmen und Agenturen hinsichtlich SEA, ist DAA certified Webanalyst, Top-Beitragender in der AdWords Community und Mitveranstalter des SEAcamp. Websites mit <u>Crawlern\*</u> strukturell erfassen Eine gut überlegte Kampagnenstruktur ist einer der Schlüsselfaktoren zum Erfolg im SEA. Zum einen sollten Anzeigengruppen möglichst fein gegliedert sein, um eine bestmögliche Relevanz zu erreichen. Zum anderen sollte die Struktur effektives Monitoring, Reporting und Management ermöglichen.

Beim Set-up, bei der Neustrukturierung bzw. beim Ausbau eines AdWords-Kontos steht somit oft die Frage der Struktur. Eine bewährte Methode ist, die Kampagnen anhand der Website zu strukturieren, d. h. Kampagnen pro Kategorie und Anzeigengruppen pro Unterkategorie und/oder Produkten zu erstellen. Genau an diesem Punkt können SEO-Tools eine große Erleichterung darstellen, indem die Struktur einer Website in eine mit Excel verarbeitbare Form gebracht wird.

Crawler wie Xenu (Kostenlos unter *http:// einfach.st/xenu*) oder der Screaming Frog (Bis 500 URL kostenlos nutzbar, *http://einfach.st/ frog*) können zur strukturellen Erfassung von Websites unterstützend eingesetzt werden. Beide Tools gibt es als kostenlose Versionen zum Download – Screaming Frog ist in der kostenlosen Version etwas limitiert.

Was genau machen diese Tools? Ein Crawler "klickt" quasi jeden Link innerhalb einer Web-

Xenu's starting point	Options	<b>X</b>
What address do you want to check? (Enter your "root URL", e.g. http://www.host.com/~user/ and all local pages linked from it will also be checked) Or click: "Browse" to check your local files the same way!	Basic Advanced	
http://www.inbiz.de/	Parallel Threads	
Check external links	1 - 🤁	Apply to all jobs
Include / Exclude: Consider URLs beginning with this as 'internal': Add Delete Inactive Do not check any URLs beginning with this: Add Delete http://www.inbiz.de/wp-content/	Current: 4 100 - Maximum depth: 2 Ask for password or certificate when needed Treat redirections as errors	Report Broken links, ordered by links Broken links, ordered by page Broken local links Redirected URLs Pap and gopher URLs Valid text URLs Site Map Statistics Orphan files
More options e-mail OK Cancel	ОК	Abbrechen

Abb. 1 + 2: Grundeinstellungen von Xenu

site an, speichert den Link und liest diverse standardisierte Informationen (z. B. Title, Meta-Informationen) aus den gewonnenen URLs aus.

Am Beispiel von Xenu lassen sich unter anderem die Crawling-Tiefe (wie viele Ebenen von der angegebenen Start-URL sollen gecrawlt werden) und die maximale Anzahl paralleler Anfragen voreinstellen (s. Abb. 01). Da auch für das hier gezeigte Anwendungsbeispiel irrelevante Bilder, CSS und JavaScript-Links ausgelesen werden, sollten diese von vornherein ausgeschlossen werden (s. Abb. 02).

Als Ergebnis liefert Xenu eine Liste im TSV-Format (TSV = Tab-Separated Values; mit Excel zu öffnen), die gefiltert auf den Type = text/html die Liste aller zu crawlenden URLs ergibt (s. Abb. O3).

Ist die Website URL-seitig in Verzeichnissen strukturiert, kann diese Liste zur Strukturierung eines AdWords-Kontos genutzt werden, indem die URLs in ihre Bestandteile aufgesplittet werden (Beispiel s.Abb. O4).

In unserem Beispiel soll aus den ersten beiden Kategorieebenen jeweils eine Kategoriekampagne entstehen. AdGroups sollen die Unterkategorien werden, die keine weiteren Unterkategorien haben. Für diese wiederum sollen eigene Kampagnen angelegt werden. Als AdGroup soll die jeweils letzte Ebene fungieren. Hauptkeyword der so

(a) - (*	(~) <del>;</del>		Beisp	siel Dump - In	Biz.tsv - Micro	soft Excel			- 0 ×
Start E	infügen Seitenlayout Formeln	Daten Überp	rüfen Ansicht	Entwicklert	ools SeoTo	ols			0 _ = x
Findfilden		= = >-	Standard	- 00, 0, 0	Bedingte	Als Tabelle Zellenformative	Gra Einfügen *	Σ · ŽT if	)
- V				0 .65 +,0	Formatierung *	formatieren *	Format *	2* und Filtern * Auswä	alen *
Zwischenablage 19	Schriftart	Ausrichtung	G Zahl			Formatvorlagen	Zellen	Bearbeiten	
A2	• (* fx http://www	.inbiz.de/							8
	А	В	C D	E					
1 Address		🔹 Status-Co💌 Stat	tus-Te💌 Type	🛒 Size	Title				
2 http://www.in	nbiz.de/	200 ok	text/htm	nl 4	47069 InBiz O	nline Marketing - SEM +.	AdWords Beratung		
47 http://www.in	nbiz.de/angebot/	200 ok	text/htm	nl S	54964 Angeb	ot und Dienstleistungen	von InBiz Online Mark	eting	
49 http://www.in	biz.de/suchmaschinenmarketing-	be 200 ok	text/htm	nl S	54993 Suchm	aschinenmarketing + SEA	Beratung - profession	nell und nachhaltig mit I	1Biz
50 http://www.in	biz.de/seo-consulting/	200 ok	text/htm	nl s	55292 SEO Co	nsulting - strukturiert un	d strategisch mit InBi		
51 http://www.ir	biz.de/retargeting/	200 ok	text/htm	nl S	55215 Retarge	eting & Behavioral 1	fargeting – Display z	ielgenau & treffsich	ier   InBiz - Online M
52 http://www.ir	biz.de/webanalyse-beratung/	200 ok	text/htm	nl S	54916 Weban	alyse Beratung - fundier	t und detailliert mit Ir	Biz	
55 http://www.in	biz.de/inbiz-das-unternehmen/	200 ok	text/htm	nl S	54265 InBiz â	€" das Unternehmen stel	It sich vor		
58 http://www.in	biz.de/thomas-gruebel/	200 ok	text/htm	nl S	53897 Thoma	s Grübel - SEA-Experte	& Webanalyst		
59 http://www.in	biz.de/marco-brenn/	200 ok	text/htm	nl s	53853 Marco	Brenn - Prokurist von InB	liz		
61 http://www.ir	nbiz.de/sea-kompendium/	200 ok	text/htm	nl 4	49988 SEA - K	ompendium Archives   I	nBiz - Online Marketir	g	
63 http://www.in	nbiz.de/sea-kompendium/dynami	-s 200 ok	text/htm	nl S	54406 Dynam	ic Search Ads (Dynamisc	he Suchanzeigen) - Se	tup, Auswertung und Op	timierung
64 http://www.in	nbiz.de/sea-kompendium/anzeige	nt 200 ok	text/htm	nl S	51405 Anzeig	entexte im SEM optimie	ren - Tipps und Tricks		
65 http://www.in	nbiz.de/sea-kompendium/kampag	ne 200 ok	text/htm	nl S	51807 Kampa	gnen- und AdGroup Stru	ktur optimieren - die I	Basis fþr einen guten O	ualitĤtsfaktor
66 http://www.in	biz.de/sea-kompendium/qualita	ts 200 ok	text/htm	nl S	52168 AdWor	ds QualitĤtsfaktor opti	mieren - Herausforde	ungen, Tipps und Tricks	
67 http://www.in	biz.de/sea-kompendium/bestand	te 200 ok	text/htm	nl S	52643 Bestan	dteile des QualitĤtsfak	tors   InBiz - Online M	arketing	
68 http://www.in	biz.de/sea-kompendium/einleitu	ng 200 ok	text/htm	nl S	51638 SEA - K	ompendium von InBiz - E	inleitung und Über	blick	
69 http://www.in	biz.de/online-marketing-glossar/	200 ok	text/htm	nl 4	46369 Online	Marketing Glossar Archi	ves   InBiz - Online Ma	arketing	
70 http://www.in	biz.de/download/	200 ok	text/htm	nl (	53601 Downle	oads - unsere Vorträge	und Blackpaper   InBi	z - Online Marketing	
83 http://www.in	biz.de/impressum-datenschutz/	200 ok	text/htm	nl S	57211 Impres	sum von InBiz Online Ma	arketing		
84 http://www.in	biz.de/kontakt/	200 ok	text/htm	nl 4	46075 Kontak	tformular InBiz Online N	larketing		
85 http://www.in	biz.de/newsletter/	200 ok	text/htm	nl 4	44363 Newsle	etter - Sei immer up-to-d	late   InBiz - Online M	arketing	
86 http://www.in	nbiz.de/sitemap/	200 ok	text/htm	nl 4	46216 Sitema	p von InBiz Online Marke	eting		
102 http://www.in	biz.de/online-marketing-glossar/	s/s 200 ok	text/htm	nl 4	43993 Suchbe	griffen			
103 http://www.in	biz.de/online-marketing-glossar/	z/: 200 ok	text/htm	nl 4	44056 Zielsei	ten			
104 http://www.in	biz.de/sea-kompendium/kampag	ne 200 ok	text/htm	nl S	53164 Kampa	gneneinstellungen			
106 http://www.in	biz.de/online-marketing-glossar/	b/ 200 ok	text/htm	nl 4	49419 Bidmar	nagement			
108 http://www.in	biz.de/online-marketing-glossar/	t/t 200 ok	text/htm	nl 4	44210 Anzeig	entitel			
110 http://www.in	nbiz.de/sea-kompendium/negativ	e-I 200 ok	text/htm	nl S	51294 Negati	ve Keyword			
111 http://www.ir	biz.de/sea-kompendium/der-anz	eiį 200 ok	text/htm	nl S	51337 Anzeig	entexte			
112 http://www.ir	biz.de/sea-kompendium/adword	5-€ 200 ok	text/htm	nl S	50648 <span o<="" td=""><td>:lass="meta-nav"&gt;←«</td><th>:/span&gt;Zurück</th><td></td><td></td></span>	:lass="meta-nav">←«	:/span>Zurück		
113 http://www.ir	biz.de/sea-kompendium/die-ziel	se 200 ok	text/htm	nl S	53340 Die Zie	lseite und der QualitĤt	sfaktor		
114 http://www.in	nbiz.de/sea-kompendium/der-qua	lit 200 ok	text/htm	nl S	52396 AdWor	ds Quality Score â€" der	QualitĤtsfaktor von	Google	
115 http://www.in	nbiz.de/sea-kompendium/anrufer	ve 200 ok	text/htm	nl (	55997 Anrufe	rweiterungen und Click	to Call		

Abb. 3: Ergebnis Crawl einer Website mittels Xenu



Abb. 4: Beispiel zur Verwendung der URL-Struktur zum Kampagnen-Set-up

gewonnenen AdGroup soll ein Broad Match Modifier Keyword, bestehend aus den Hauptkategorien und dem spezifischen Unterkategorienamen, werden (Hinweis: Mit dem "Modifierzeichen "+" können Keywords mit der Einstellung "Weitgehend passend", also Broad Match, genauer ausgesteuert werden – weitere Infos unter *http://einfach.st/ modif1*).

### **ONLINE MARKETING » SEO -> SEA**

Α	В	С	D	E	F	G	Н	1
Address		Address						
https://www.otto.de/sport/herren/		https:		www.otto.de	sport	herren		
https://www.otto.de/sport/herren/accessoires/	Textk	onvertierungs-Assistent - Schritt 2	von 3	2 ×	sport	herren	accessoires	
https://www.otto.de/sport/herren/bademoden/	Diese	s Dialogfeld ermöglicht es Ihnen, Trer ierten Daten sehen, wie Ihr Text ersci	nzeichen festzulegen. neinen wird.	Sie können in der Vorschau der	sport	herren	bademoden	
https://www.otto.de/sport/herren/funktionswaesche/	Tree	nzeichen Tabstono			sport	herren	funktionswaesche	
https://www.otto.de/sport/herren/hemden/		Semikolon 📄 Aufeinanderf Komma	olgende Trennzeichen	als ein Zeichen behandeln	sport	herren	hemden	
https://www.otto.de/sport/herren/hosen/		Leerzeichen Texterkennungs	reichen:		sport	herren	hosen	
https://www.otto.de/sport/herren/jacken/	Dat	envorsghau			sport	herren	jacken	
https://www.otto.de/sport/herren/parkas/					sport	herren	parkas	
https://www.otto.de/sport/herren/pullover/	60 65 65	tps: www.otto.de spo tps: www.otto.de spo	rt herren acces	soires	sport	herren	pullover	
https://www.otto.de/sport/herren/shirts/	ht ht	tps: www.otto.de spo	rt herren badem rt herren funkt:	oden ionswaesche +	sport	herren	shirts	
https://www.otto.de/sport/herren/shorts-bermudas/		Abbr	echen < Zurück	: Weiter > Fertig stellen	sport	herren	shorts-bermudas	
https://www.otto.de/sport/herren/sneaker/		nups:	1	www.otto.de	sport	herren	sneaker	
https://www.otto.de/sport/herren/sportschuhe/		https:		www.otto.de	sport	herren	sportschuhe	
https://www.otto.de/sport/herren/sporttaschen/		https:		www.otto.de	sport	herren	sporttaschen	
https://www.otto.de/sport/herren/jacken/trainingsjacken/		https:		www.otto.de	sport	herren	jacken	trainingsjacken
https://www.otto.de/sport/herren/sportschuhe/laufschuhe/		https:		www.otto.de	sport	herren	sportschuhe	laufschuhe
https://www.otto.de/sport/herren/sportschuhe/trainingsschuhe/		https:		www.otto.de	sport	herren	sportschuhe	trainingsschuhe
https://www.otto.de/sport/herren/jacken/softshelljacken/		https:		www.otto.de	sport	herren	jacken	softshelljacken

Abb. 5: Beispiel aufgesplitteter URLs

Der Ablauf, um hieraus eine in den AdWords-Editor (den Offline Editor für AdWords von Google, kostenlos unter *http://einfach.st/adwedit*) ladbare Datenkopie zu bekommen, ist folgender:

- » URLs in Verzeichnis- und Dateinamen aufsplitten.
- » Ermitteln, ob die Kategorie Unterkategorien hat.
- » Entscheidung bzgl. Kampagnenname und Generierung Kampagnenname.
- » AdGroup-Namen generieren.
- » Keyword generieren.

Das Aufsplitten der URLs kann durch die Excel-Funktion "Text in Spalten" mit Trennzeichen "/" (s. Abb. 05) oder durch die Funktion TEIL in Verbindung mit FINDEN erfolgen.

Verkürzt auf die wirklich relevanten Kategoriespalten (Spalten F bis I aus Abbildung 5) kommt im Ergebnis eine Liste mit vier Spalten heraus. Zunächst wird nun ermittelt, auf welcher Hierarchieebene die jeweilige Unterseite aufgehangen ist (Spalte E):

## =WENN(D4<>"";4;WENN(C4<>"";3))

Als Nächstes soll ermittelt werden, ob eine Kategorie Unterseiten hat, um die Zuordnung zur entsprechenden Kampagne treffen zu können. Dies erfolgt durch die Benutzung der ZÄHLENWENN-Funktion, welche prüft, wie oft die jeweilige Kategorie in der Liste vorkommt (Spalte F):

=WENN(ZÄHLENWENN(C:C;C4)<>1;"Ja";"Nein")

Basierend auf dieser Information wird entschieden, wie der Kampagnenname zusammengesetzt wird (Spalte G):

="01 - "&GROSS2(A4)&" - "&GROSS2(B4)&WENN(F4= "Ja";" - "&GROSS2(C4);"") Umlaute können hierbei zudem durch die WECHSELN-Funktion übersetzt und Großschreibung durch die GROSS2-Funktion umgesetzt werden.

Nach ähnlichem Muster werden Anzeigengruppennamen (Spalte H) generiert:

=WENN(E4=3;GROSS2(C4);GROSS2(D4))

Last, but not least, wird das Hauptkeyword (Spalte I)der AdGroup hervorgebracht:

```
="+"&GROSS2(A4)&" +"&GROSS2(B4)&" +"&WENN(E4=3;
GROSS2(C4);GROSS2(D4))
```

Das Ergebnis lässt sich nun per Copy & Paste in den AdWords-Editor einfügen. Danach sind noch Kampagneneinstellungen, AdGroup-Gebot und Matchtypes zu setzen – fertig ist eine Rumpfkampagne.

Bonustipp: Nach ähnlichem Prinzip können auch Template-basierte Anzeigentexte oder AdGroup-spezifische Sitelinks erstellt werden.

# Website-Inhalte mittels Scraping-Techniken ermitteln

Ist eine Website nicht in Verzeichnisse strukturiert, können wir, basierend auf diesen URLs, mittels Scraping-Techniken nutzbare Informationen gewinnen. Im Folgenden werden hierzu die SEO-Tools for Excel (siehe hierzu ausführlich den Titelbeitrag der Website Boosting Ausgabe 20) genutzt. Dieses kostenlose Excel-Add-in bietet viele nützliche Funktionen, z. B. Scraping/XPathOnUrl. Mittels dieser Funktion können, basierend auf einer URL und dem sogenannten XPath, ganz dediziert Einzelelemente einer Webseite (z. B. Preis, Marke) ausgelesen werden. Voraussetzung für die

						· · · · · · · · · · · · · · · · · · ·		
sport	herren	jacken		з	Ja	01 - Sport - Herren - Jacken	Jacken	+Sport +Herren +Jacken
sport	herren	parkas		З	Nein	01 - Sport - Herren	Parkas	+Sport +Herren +Parkas
sport	herren	pullover		з	Nein	01 - Sport - Herren	Pullover	+Sport +Herren +Pullover
sport 🚽	horron	chirte		з	Nein	01 - Sport - Herren	Shirts	+Sport +Herren +Shirts
sport	⊿ 🍅 01 - Spor	rt - Herren		3	Nein	01 - Sport - Herren	Shorts-Bermudas	+Sport +Herren +Shorts-Bermudas
sport	Access	oires		3	Nein	01 - Sport - Herren	Sneaker	+Sport +Herren +Sneaker
sport	Gaden	ongwoorsho	uhe	4	Ja	01 - Sport - Herren - Sportschuhe	Laufschuhe	+Sport +Herren +Laufschuhe
sport	Hend	an	gsschuhe	4	Ja	01 - Sport - Herren - Sportschuhe	Trainingsschuhe	+Sport +Herren +Trainingsschuhe
sport	Hosen			3	Ja	01 - Sport - Herren - Sportschuhe	Sportschuhe	+Sport +Herren +Sportschuhe
sport	Parkas			3	Nein	01 - Sport - Herren	Sporttaschen	+Sport +Herren +Sporttaschen
sport	<b>Pullove</b>	er		3	Nein	01 - Sport - Herren	Strickjacken	+Sport +Herren +Strickjacken
sport	i Shirts			3	Nein	01 - Sport - Herren	Sweatjacken	+Sport +Herren +Sweatjacken
sport	Chorts 👰	-Bermudas		3	Nein	01 - Sport - Herren	Sweatshirts	+Sport +Herren +Sweatshirts
sport	<b>Sneake</b>	er			Nein	01 - Sport - Herren	Trainingsanzuege	+Sport +Herren +Trainingsanzuege
sport	Sportt	aschen			Nein	01 - Sport - Herren	Trikots	+Sport +Herren +Trikots
sport	Stricky	acken			Nein	01 - Sport - Herren	Westen	+Sport +Herren +Westen
	Sweat	acken						- 1
	Trainin							
		s sanzacyc						
	Weste	n						
	⊿ 200 01 - Spor	rt - Herren - Jacken						
	Jacken	1						
	Coftsh 👔	elljacken						
	🥡 Trainin	gsjacken						
	a 🚵 01 - Spor	rt - Herren - Sportschul	he					
	and a second	• •	-					

Abb. 6: Fertiger Dump zum Copy & Paste in den AdWords-Editor

Anwendung ist eine valide HTML-Struktur der Website. Daneben bietet das Tool vorgefertigte Funktionen für die Ermittlung bestimmter Standard-HTML-Elemente (z. B. Title, H1-Überschrift) oder zum Zählen bestimmter Zeichenketten (IsFoundOnPage).

Um die Funktion XPathOnUrl nutzen zu können, muss der XPath – also die genaue Adresse – eines Elementes auf der Seite bekannt sei. Dazu stellen Browser Hilfsmittel zur Verfügung.

Im Chrome kann der XPath mit Klick der rechten Maustaste auf ein zu analysierendes Element, die Auswahl "Element untersuchen" und nach Öffnung der Konsole im unteren Bereich mit nochmaligem Klick auf die rechte Maustaste + "Copy XPath" ermittelt werden.

Beim Firefox funktioniert das mit installiertem Plug-in Firebug ähnlich: Mit rechter Maustaste auf ein Element klicken, Element mit Firebug untersuchen, in der sich öffnenden Konsole rechte Maustaste auf das Element und XPath kopieren.

Als einleitendes Beispiel folgt hier die Ermittlung des Kategoriebaumes einer Seite über den Breadcrump, der wie oben gezeigt zur Erstellung einer Kampagnengrundstruktur verwendet werden kann. Der Wert des XPathes (Zelle B2) wird, wie in Abb. 07 und 08 gezeigt, im Browser ermittelt. (siehe Abbildung 9)

Im Folgenden sollen drei konkrete Anwendungsbeispiele aus dem SEA-Bereich zum Thema Scraping im Detail erläutert werden:

# 1. Fehlerhafte Seiten mittels Scraping aufspüren

Im SEA gebuchte Zielseiten müssen technisch funktionieren, bei Shops Produkte beinhalten (und keine Fehlermeldungen) und eine möglichst hohe Relevanz zur Suchanfrage haben. Alle drei Punkte lassen sich mit der Kombination Xenu und SEO-Tools for Excel prüfen.

okies• 🖍 CSS• 🗋 Formulare• 📾 Bilder• 🕦 Informationen• 📄 Sonstiges• 🥢 Kontur• 🖉 Größe ändern• 💥 V HTML kopieren NPath kopieren CSS-Pfad konf Den XPath des Elements in die Zwischenablage kopieren Paste HTML	
HTML kopieren InnerHTML kopieren XPath kopieren CSS-Pfad kon[Den XPath des Elements in die Zwischenablage kopieren Paste HTML	onstiges+ 🥖 Kontur+ 🥒 Größe ändern+ 💥 We
HTML kopieren InnerHTML kopieren XPath kopieren CSS-Pfäd kof Den XPath des Elements in die Zwischenablage kopieren Paste HTML	
XPath kopieren CSS-Pfad kof <mark>i</mark> Den XPath des Elements in die Zwischenablage kopieren Paste HTML	
	h des Elements in die Zwischenablage kopieren
Ereignisse protokollieren	1
Zum Sichtbereich scrollen Suche	en Suche
Neues Attribut	
HTML bearbeiten	ft. mit
Konsole HTML - CSS Skript DOM Element löschen Del	Del
div.snippet < div < div#snipppet_set < Alle auf-/zuklappen * ml	* ml
# <style type="text/css"></style>	



Abb. 7 + 8: XPath mit Chrome und Firefox ermitteln

Die erste Stufe der technischen Funktionstüchtigkeit (eigentlich kein Scraping) kann geprüft werden, indem man eine Linkliste in Xenu lädt und prüfen lässt oder die Liste der URLs mittels der Excel-Funktion aus den SEO-Tools for Excel "HttpStatus" prüft. Die Funktion gibt zudem bei Redirects das Redirect-Ziel zurück, welches wiederum getestet werden kann. Tracking-Parameter beziehungsweise Redirects können zuvor durch die WECHSELN-Funktion entfernt und decodiert werden (siehe Abbildung 10).

Eine weitere Fehlerquelle hinsichtlich der Zielseiten ist die Verlinkung in Kategorien oder in Suchergebnisse, zu denen es aufgrund von Sortimentsveränderungen keine Produkte mehr gibt. Liefern diese Seiten keinen technischen Fehler, können sie inhaltlich hinsichtlich bestimmter Fehlertexte oder der Anzahl bestimmter Elemente geprüft werden.

Fehlermeldungen (z. B. keine Produkte bei Verlinkung in ein Suchergebnis oder in eine Kategorie) sind mit der Funktion IsFoundOnPage zu ermitteln (siehe Abbildung 11).

# 2. Relevanz von Zielseiten mittels Scraping bewerten

Beim Thema Relevanz kann man weit ausholen und beispielsweise ein eigenes Bewertungssystem mit gewichteten Faktoren zusammenstellen. Bei Shops und generischen – nicht produktbezogenen - Kampagnen kann auch schon die Anzahl der auf einer Zielseite vorhandenen Produkte eine Aussage über die Relevanz der Seite treffen. Produkte sind im HTML-Code einer Seite meist anhand bestimmter CSS-Klassen erkennbar (z. B. Product-Name). Wenn man wie oben schon beschrieben ein Produktelement (z. B. den Namen oder das Bild) auswählt und mit der rechten Maustaste das Element untersucht, wird man in der dann aufgehenden Konsole schnell

	B3	✓ ( SYPATHONURI(B1;B2)				
	А	B				
1	URL	http://www.zalando.de/tamaris-sandalette-black-silver-ta111i002-q11.html				
2	Xpath	//*[@id="content"]/div/div[1]/ul				
3	Ergebnis	Startseite > > Schuhe > Sandalen > Riemchensandalen > Sandale - black/silver				
	1800113	Starbeiter im Schaner Sandalen internenensandalen Sandale Stadigsmer				

Abb. 9: Scraping des Kategoriebaumes mittels XPathOnURL

	B3	<del>-</del> (9	$f_{x}$	=Http	oStatus(A3)	
	4	Α		_		В
1	http://www.ir	ibiz.de/			200 OK	
2	2 http://www.inbiz.de/FEHLER				404 Not Fou	nd
3	3 http://www.inbiz.de/downloads		301 Moved I	Permanently http://www.inbiz.de/download/		

Abb. 10: Linkcheck mit HttpStatus

	C4 🔻 💿	<pre>f<sub>★</sub> =IsFoundOnPage(A4;B4)</pre>			
		А		В	С
1	URL			Fehlertext	Anzahl Produkte
2	http://www.zalando.de/g	esicht-reinigung/yes-to-blue	berries/	Noch nicht aufgeben	0
3	http://www.zalando.de/c	lamenbekleidung-jeans/		Noch nicht aufgeben	0
4	http://www.zalando.de/s	ports-outdoorschuhe-damen	/groesse-35/	Noch nicht aufgeben	1

Abb. 11: Scraping zum Check auf Fehlermeldung



Abb. 12: Anzahl der Produkte auf einer Seite durch Zählen ermitteln

	C3	• (0	ƒ <sub>≪</sub> =XPathOnUrl(A3;B3)		
		А		В	С
1	URL			XPath	Anzahl
2	http://www.fal	nrrad-xxl.de/fal	hrraeder/bmx	/html/body/div/div/div[2]/div/div[3]/div/p/strong	28
3	http://www.fal	nrrad-xxl.de/fal	hrraeder/mountainbikes	/html/body/div/div/div[2]/div/div[3]/div/p/strong	492

Abb. 13: Anzahl Produkte direkt auslesen



Abb. 14: Prüfung, ob Keyword in H1 und Anzahl

solche CSS-Elemente oder IDs finden. Diese können mithilfe der Funktion IsFoundOnPage gezählt werden (siehe Abbildung 12).

Zudem wird die Anzahl der vorhandenen Produkte häufig direkt auf der Seite angezeigt, welche mittels der Funktion XPathOnURL ausgelesen werden kann (siehe Abbildung 13).

Dieses Vorgehen eignet sich sehr gut, um regelmäßig bestimmte Seitenkategorien hinsichtlich ihrer Produktbestückung zu prüfen.

Weiterhin spiegelt aus SEA-Sicht

eine relevante Landingpage\* erst mal die Suchanfrage des Users auf den ersten Blick wider. Der dazu auffälligste und auslesbare Punkt ist die Hauptüberschrift (in der Regel H1), welche sich mit der Funktion HtmlH1 ermitteln und prüfen lässt.

Ob der Begriff überhaupt irgendwo auf einer Seite oder wie oft (hoffentlich liest Karl Kratz das nicht ⓒ) er vorkommt, kann mittels der Funktion IsFoundOnPage ermittelt werden (Abbildung 14).

Gewichtet man diese Faktoren,

\*siehe Online-Glossar unter www.websiteboosting.com

	C6	✓ ( f <sub>*</sub> =XPathOnUrl(B\$1;\$B6;"alt")	
	А	В	с
1	URL	http://www.mediamarkt.de/mcs/product/SAMSUN	G-GT-P5200-TAB3-10.1-3G-WIFI-16-GB-goldbraun,48353,462778,995083.html
2			
3	Spalte	Xpath	
4	Name	/html/body/div[3]/div/div[2]/h1	SAMSUNG GT-P5200 TAB3 10.1 3G+WIFI 16 GB goldbraun
5	Kategorie	/html/body/div[3]/nav/ul/li[4]/a	Tablets
6	Marke	/html/body/div[3]/div/div[2]/div/div/img	SAMSUNG
7	Bild URL	/html/head/meta[9]	http://pics.redblue.de/doi/pixelboxx-mss-60542514/fee_325_225_png/SAMSU
8	Preis	/html/head/meta[12]	299.00
9	ID	//*[@id="product-details"]/div[2]/div[1]/dl/dd[1]	1749825

Abb. 15: Scraping zur Erstellung einer Produktdatenliste

kann aus einer Liste von vielen möglichen Zielseiten die theoretisch am besten passende herausgefunden werden. Dieses Vorgehen eignet sich beispielsweise zur Prüfung von Keywords mit besonders niedrigem Qualitätsfaktor und hohen Kosten.

# 3. Scraping zur Erstellung eines Produktdatendumps

Ein weiteres Anwendungsbeispiel zu Scraping-Techniken ist die Erstellung eines Produktdatendumps zur Verwendung in Google Shopping/Merchant Center. Basierend auf einer Liste von Produkt-URLs soll gezeigt werden, wie aus den in den Seiten befindlichen Informationen eine Liste von Produktinformationen zusammengestellt werden kann.

Als Pflichtfelder sollen URL, Name, Kategorie, Marke, Bild-URL, Preis und ID ermittelt werden. Sind die Informationen in der Website vorhanden, können Sie mittels XPathOnURL ausgelesen werden. Sind die Ausgangsseiten gleich strukturiert, muss der jeweilige XPath nur einmal ermittelt werden und kann auf beliebig viele Produktseiten angewendet werden.

Im Unterschied zu den bisher gescrapten Elementen besitzen manche Tags (z. B. IMG-Tag, Meta-Tags) keinen eigenen Inhalt. Die Quelle des Bildes ist in einem Attribut (src = Quelle) abgelegt, die Inhalte der Meta-Tags im Attribut content. Hierzu bietet die Funktion XPathOnURL zusätzliche optionale Parameter, mit deren Hilfe auch auf Attribute eines HTML-Elementes zugegriffen werden kann. Am in Abb. 15 gezeigten Beispiel trifft das auf die Marke zu, die aus dem ALT-Tag des Marken-Logos extrahiert wird, sowie auf die Bild-URL und den Preis (bezieht sich auf Attribut content der Meta-Informationen). Die so gewonnene Liste aus XPathes kann für weitere URLs zur Vorbereitung des Produktdumps genutzt werden. Der Übersichtlichkeit halber werden die Daten zeilenweise dargestellt. Zur Erstellung eines Dumps empfiehlt sich eine Anordnung in Spalten.

### Fazit

SEO-Tools wie Crawler und Scraping-Tools können einen SEA-Kampagnen-Manager bei der Arbeit unterstützen und zeitaufwendige Tasks insbesondere durch Skalierbarkeit deutlich erleichtern bzw. verkürzen.

Weitere Anwendungsbeispiele sind:

- » Erstellung von Anzeigentexten teilautomatisch beziehungsweise Template-basiert
- » Keyword-Recherche (z. B. Auslesen von Meta-Keywords, sofern vorhanden)
- » Konkurrenzanalyse

Limitiert sind die SEO-Tools for Excel hinsichtlich der Performance. Listen von mehr als 5.000 URLs können je nach Hardware, Ladezeit der zu prüfenden Seiten und Seitengröße schnell zu Problemen und Abbrüchen führen. Mit guter Vorsortierung (z. B. absteigend nach Anzahl der Klicks bei Zielseitenchecks) und sequenzieller Abarbeitung kann dieses Problem oft umgangen werden.

Die mit den vorgestellten Browsermethoden ermittelten XPathes funktionieren mit den SEO-Tools for Excel leider nicht immer. Mitunter hilft es, einen anderen Browser zu benutzen, mal liegt es aber auch an der Seitenstruktur. Häufig hilft auch, wenn man nach dem Inhalt des zu ermittelnden Elementes sucht und es an anderer, nicht direkt sichtbarer und verwendbarer Stelle wiederfindet (bspw. in den Meta-Informationen).

Bei allen Schritten sollte bedacht werden, dass die Ergebnisse und deren Qualität nur so gut sind wie die dahinterliegenden Konzepte. Viele der vorgestellten Tasks (insb. die Erstellung von Rumpfkampagnen) erfordern manuelle Nacharbeit und Prüfung durch den Kampagnenmanager. SEA ist und bleibt Handwerk, darum gilt wie so oft die Regel: Lassen Sie sich die Arbeit abnehmen, aber nicht das Denken!

Alle gezeigten Beispiele gibt es zum Download unter *http://einfach.st/ inbiz*¶

### TIPP

Beim Scrapen fremder Webseiten sollte man immer darauf achten, nicht zu viele automatisierte Anfragen gleichzeitig zu erzeugen. Bekommt der Sitebetreiber durch solche Abfragen ein Performanceproblem oder gar messbare Timeouts auf dem Webserver, könnte man solche Abfragen ggf. als sog. Denial-of-Service(DoS)-Attacke interpretieren, die strafrechtlich durchaus verfolgbar wäre. Normalerweise machen derartige Aktivitäten keine nennenswerten Probleme und bleiben sogar in der Regel völlig unbemerkt, weil die meisten Webanalytics-Tools maschinelle Anfragen gar nicht auswerten bzw. mit ausgeben. In den Logfiles des Webservers sind sie allerdings exakt mit Datum und Uhrzeit sowie mit IP-Adresse gespeichert und können im Fall eines Schadens durch offiziell eingeschaltete Stellen durchaus rückverfolgt werden. Bei allzu intensivem Scraping innerhalb einer sehr kurzen Zeitspanne sollte man sich dieser möglichen Folgen also bewusst sein – auch wenn das "Risiko" in der Praxis überschaubar gering ausfällt und daher meist einfach ignoriert wird.