

### Optimierung der Webseiten von Forschungsdaten- und Datenservicezentren für den Zugriff durch Suchmaschinen: AG Metadaten des Ständigen Ausschusses FDI des RatSWD

Weichselgartner, Erich

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

#### Empfohlene Zitierung / Suggested Citation:

Weichselgartner, E. (2014). *Optimierung der Webseiten von Forschungsdaten- und Datenservicezentren für den Zugriff durch Suchmaschinen: AG Metadaten des Ständigen Ausschusses FDI des RatSWD*. (RatSWD Working Paper Series, 234). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-75374-2>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

Weichselgartner, Erich

**Working Paper**

Optimierung der Webseiten von Forschungsdaten- und  
Datenservicezentren für den Zugriff durch Suchmaschinen: AG  
Metadaten des Ständigen Ausschusses FDI des RatSWD

RatSWD Working Paper, No. 234

**Provided in Cooperation with:**

German Data Forum (RatSWD)

*Suggested Citation:* Weichselgartner, Erich (2014) : Optimierung der Webseiten von Forschungsdaten- und Datenservicezentren für den Zugriff durch Suchmaschinen: AG Metadaten des Ständigen Ausschusses FDI des RatSWD, RatSWD Working Paper, No. 234, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin

This Version is available at:

<http://hdl.handle.net/10419/96242>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

234

## Optimierung der Webseiten von Forschungsdaten- und Daten- servicezentren für den Zugriff durch Suchmaschinen

AG Metadaten des Ständigen  
Ausschusses FDI des RatSWD

Erich Weichselgartner

April 2014



## Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

---

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

*RatSWD Working Papers* sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienenen empirischen Forschungsarbeiten sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

# Optimierung der Webseiten von Forschungsdaten- und Datenservicezentren für den Zugriff durch Suchmaschinen

Erich Weichselgartner

*Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID),  
54296 Trier*

Der Wissenschaftsrat tritt für maximale Publizität digitaler Forschungsdaten ein. Allerdings sei die Zugänglichkeit von Forschungsdaten nach wie vor häufig nicht oder nicht hinreichend gegeben (WR Drs. 2359-12). In den Fällen, in denen die Zugänglichkeit gegeben ist, sollten sie möglichst leicht gefunden werden. Weil auch im wissenschaftlichen Kontext gerne Suchmaschinen für die Informationsbeschaffung eingesetzt werden, ist ein Weg, die Zugänglichkeit zu verbessern, die Optimierung der Webseiten der Forschungsdatenanbieter für den Zugriff durch Suchmaschinen. Da sich die verbreiteten Suchmaschinen in wesentlichen Grundprinzipien nicht unterscheiden (Einsammeln der Inhalte von Webseiten mit Hilfe sogenannter Crawler, Aufbau eines Index, Verarbeitung der Suchanfragen, Anzeige der Trefferlisten), führen die Optimierungshinweise unabhängig davon zum Erfolg, welche konkrete Suchmaschine die Nutzerinnen und Nutzer bevorzugen. Die folgenden Hinweise beruhen speziell auch auf Erfahrungen, die das ZPID mit dem Einsatz der Psychologie-Suchmaschine PsychSpider bei einer Reihe von Datenzentren gesammelt hat.

## Einleitung

Handreichungen darüber, wie man Websites für Suchmaschinen optimiert, sind zuhauf im Internet zu finden (Bing 2013; Google 2011; Suchmaschinentricks 2013; Yandex 2013), sogar für den akademischen Bereich (Beel, Gipp & Wilde, 2010). Dabei steht häufig der Marketing-Aspekt im Vordergrund, nämlich wie man das eigene Angebot mit allen Mitteln auf die obersten Plätze von Trefferlisten katapultiert (also das Ranking des eigenen Angebots verbessert). Dies mag ein Nebeneffekt der nachfolgenden „Best Practice“-Tipps aus dem Bereich Suchmaschinenoptimierung (SEO, engl.: Search Engine Optimization) sein, im Vordergrund steht aber, die Websites der Datenservicezentren (DSZ) und Forschungsdatenzentren (FDZ) besser zu erschließen, d.h. sie sollen nicht nur besser gefunden werden, sondern man soll sich auf ihnen auch besser zurechtfinden können – egal ob der Nutzer ein Mensch oder eine Maschine ist. Ziel ist beispielsweise, möglichst einfach von der Information — dass ein Datenzentrum über einen bestimmten Datensatz verfügt — zur Beschreibung des Datensatzes und schließlich zum Datensatz selbst zu gelangen. Die nachfolgenden Hinweise gelten grundsätzlich für alle Inhalte von Websites, und damit auch für Forschungsdaten, letzteren ist dennoch ein eigener Abschnitt gewidmet („8. Forschungsdaten“), in dem bisher gemachte Erfahrungen adressiert werden.

Die nachfolgenden Hinweise beziehen sich auf **Websites**, also Web-Auftritte oder Web-Präsenzen, deren Inhalte üblicherweise in HTML geschriebene **Webseiten** sind. Solche HTML-Dokumente müssen von Web-Browsern interpretiert werden, damit sie von Menschen (am Bildschirm) intuitiv lesbar sind. Eine Suchmaschine sieht eine Webseite allerdings nicht so, wie sie ein moderner Web-Browser am Bildschirm darstellt, sondern wie ein – in den Anfangszeiten des Web verwendeter – textbasierter Browser (z.B. Line Mode Browser oder Lynx).

### Vorab die wichtigsten Tipps zusammengefasst

- Die Datei robots.txt anlegen
- Eine Sitemap anlegen
- Saubere HTML-Syntax verwenden (Suchmaschinen „verstehen“ die Tags)
- Meta-Elemente wie Titel, Schlagwörter und Beschreibung eintragen
- Verständliche und bedeutsame URLs verwenden
- Inhalte so verfassen, dass sie auch mittels textbasiertem Browser lesbar sind
- Unnötige Multimedia-Elemente (Audio, Video) und unnötiges Scripting (Flash, JavaScript) vermeiden
- Zu Verweisen (Links) immer Ankertexte angeben
- Gutes Hypermedia-Design anwenden; wichtige Inhalte stets wechselseitig verlinken (z.B. Beschreibung eines Datensatzes und Datensatz selbst)
- Semantische Auszeichnungen vornehmen (Microdata)

# 1. Kopfbereich (Header)

## Titel (<title>)

Seitentitel durch *Titel-Tag* (<title>) im *Header* (<head>-Tag im HTML-Code) vergeben. **Der Titel erscheint in der Trefferliste!**

- Titel möglichst kurz, prägnant und selbsterklärend formulieren
- Unterschiedliche Titel für jede (Unter-)Seite erstellen

Beispiel<sup>1</sup>:

```
<title> Methodenbericht 2011</title>
<title> Deutscher Alterssurvey, 2. Welle 2002</title>
```

## Meta-Elemente

Die Metadaten werden ebenfalls im Kopf-Bereich eines HTML-Dokuments, also im head-Element, eingefügt. **Suchmaschinen werten die Meta-Elemente aus!**

### Beschreibung (description)

- Hier sollte eine aussagekräftige, konkrete und somit individuelle Beschreibung des Inhalts der Seite stehen, die nicht künstlich mit Keywords überfrachtet wird.
- **Die Beschreibung wird als Snippet („Schnipsel“) in der Trefferliste angezeigt.**

Beispiel:

```
<meta name="description" content="FRAGEBOGEN_NUMMER_88-3-
STUDIE_10.994">
<meta name="description" content="Demographic Differences in
Life Course Dynamics in Eastern and Western Germany.">
```

### Schlagwörter (keywords)

- Schlagwörter erschließen den Inhalt eines Dokuments mit wenigen **suchrelevanten Begriffen**. Die Schlagwörter werden durch Kommata voneinander getrennt.
- Google bietet ein Tool an, das die Auswahl relevanter Schlagwörter erleichtert: <http://adwords.google.de/keywordplanner> (Sie benötigen dazu ein Google-Konto).

Beispiel:

```
<meta name="keywords" content="NEPS, Nationales
Bildungspanel, National Educational Panel Study">
```

---

<sup>1</sup> Alle Beispiele in diesem Dokument sind realen Webseiten entnommen. Sie beinhalten mitunter sowohl schlechte (durchgestrichene) als auch empfohlene Varianten.

## Autor

- Teilen Sie den Namen des Autors (des inhaltlich Verantwortlichen) für das Dokument mit.

Beispiel:

```
<meta name="author" content="Weichselgartner, wga@zpid.de">
```

## Anweisungen für Webcrawler

- Erlauben Sie einem Crawler ausdrücklich, Inhalte aus dem aktuellen Dokument und aus weiteren Dokumenten, zu denen Links gesetzt sind, zu besuchen und zu indexieren.

Beispiel:

```
<meta name="robots" content="all">
```

## 2. Webseiten-Struktur

### URLs verständlich benennen

- URLs sollten verständliche Namen erhalten; **URL-Namen erscheinen in den Suchergebnissen!**
- Bedeutsame Namen für URLs verwenden („speaking URL path“) statt langkettige unverständliche Buchstaben- und Ziffernabfolgen. Dies ist häufig ein Problem mit Content-Management-Systemen wie z.B. TYPO3. Zur Abhilfe gibt es spezielle Extensions:

<http://typo3.org/extensions/repository/view/cooluri>

<http://typo3.org/extensions/repository/view/realurl>

- Interpunktion verwenden (z.B. Bindestriche)

Beispiel:

<http://www.uni-trier.de/index.php?id=368>

<http://www.gesis.org/en/services/study-planning/>

### Einfache Verzeichnis-Struktur

- Die Verzeichnis-Struktur sollte einfach gehalten werden, dabei sollte eine extreme Verschachtelung vermieden werden.

Beispiel:

```
.../dir1/dir2/dir3/dir4/dir5/dir6/page.html
```



## Deduktive Seitennavigation

- Vom Allgemeinen zum Besonderen: nachvollziehbare Hierarchie herstellen, komplexe Linknetze vermeiden, ausschließlich Textlinks verwenden (keine Grafiken, Videos oder JavaScript).

## XML-Sitemap

- Bei großen Websites und bei komplizierten Verlinkungsstrukturen kann eine XML-Sitemap dem Crawling auf die Sprünge helfen (unabhängig von einer HTML-Sitemap für die menschlichen Nutzer). Ein Open-Source-Sitemap-Generator findet sich hier:

<https://code.google.com/p/googlesitemapgenerator/>

## 3. Webseiten-Gestaltung

### Text

- Der Quelltext ist für Suchmaschinen hochrelevant. Um die eigene Webpräsenz diesbezüglich beurteilen zu können, empfiehlt sich der Gebrauch eines reinen Textbrowsers (z.B. Lynx, siehe unten). Was man mit Lynx sieht, bzw. was sauber im Quelltext steht, das kann auch eine Suchmaschine gut auslesen.
- Problematisch wird es bei Bildern, Videoinhalten, Flash, JavaScript und HTML5. Während die „großen“ Suchmaschinen wie Google mit Multimedia zunehmend besser zurechtkommen, ist das bei „kleinen“ Spezial-suchmaschinen nicht der Fall.

### Bilder

- Bilder sollten über das "ALT"-Attribut mit beschreibendem Text versehen werden (wird angezeigt, wenn die Grafik nicht dargestellt werden kann); Bildunterschriften hinzufügen; Text nicht in bildhafter Form (als Grafik) darstellen (nicht maschinenlesbar und nutzerunfreundlich).

Beispiel:

```

```

### JavaScript

- <noscript>-Tag einfügen; die eingebetteten Inhalte sollten dem entsprechen, was bei aktiviertem JavaScript im Browser angezeigt wird.

### Videos

- Videos sollten im HTML-Code beschreibenden Text erhalten.

### Verborgener Text

- Keine „Tricks“ wie weiße Schriftart auf weißem Grund, um Text in ein Dokument „einzuschmuggeln“. Suchmaschinen erkennen dies als Manipulationsversuch und strafen es ab.

### Browser

- Sie sollten einen Textbrowser zur Überprüfung Ihrer Website verwenden, da diese hier so dargestellt wird, wie die Crawler sie „sehen“ (erfassen). Verfügbar sind beispielsweise der Line Mode Browser des CERN (<http://www.w3.org/LineMode/>) oder Lynx von der University of Kansas (<http://lynx.isc.org/>).

### Wichtige Informationen als maschinenlesbarer Text

- Die **relevanten** Inhalte Ihrer Webseite sollten nur maschinenlesbaren Text aufweisen.

## 4. Content-Optimierung

### Doppelte Seiten vermeiden

- „Dubletten“ werden als Manipulationsversuch (Suchmaschinen-Spamming) gewertet und abgestraft.
- Ggf. serverseitige 301-Weiterleitung einrichten (HTTP-Statuscode 301 Moved Permanently)
- Mit dem Canonical-Tag von einem oder mehreren Duplikaten einer Webseite auf das Original verweisen

Beispiel:

```
<link rel="canonical" href="http://www.example.de/richtige-seite">
```

### Überschriften verwenden

- Die Verwendung von Überschriften hat einen doppelten Vorteil: Sie erhöht die Lesbarkeit für Mensch und Maschine. Verwenden Sie dazu die HTML-heading-Tags <h1> bis <h6>.

### Aussagekräftige Ankertexte verwenden

- Der Ankertext soll sowohl Nutzern als auch Suchmaschinen prägnant etwas über die Seite mitteilen, zu der verlinkt wird. Dies ist umso wichtiger, wenn die URL „nicht sprechend“ ist (siehe Abschnitt 2).

Beispiel:

```
<a href="./?page=27&id=43">German Survey of Income and Expenditure</a>
```

### Weitere Hinweise zum Content

- Gebräuchliche Dateiformate verwenden (die Crawler müssen sie lesen können), kurze und ausdrucksstarke Dateinamen vergeben, alle Seiten mit Metadaten ausstatten (vgl. Abschnitt 1).

## 5. Semantische Technologien

Suchmaschinen können anhand von HTML-Tags zwar erkennen, dass eine Textstelle wichtig ist (z.B. weil sie mit `<strong>` hervorgehoben ist), aber nicht warum sie wichtig ist. Letzteres tangiert die Domäne der Semantik (Bedeutung). Gattani (2010) formuliert es so: „Semantic web is about teaching machines to read web pages, which are designed to be read by humans“. Ein Ansatz im Rahmen von semantischen Technologien, Inhalte maschinenlesbar zu machen und für die Crawler von Suchmaschinen zu markieren, sind Microdata und `schema.org`. Trotz anderer Konzepte wie Linked Open Data ist der Ansatz im Zusammenhang mit der Websuche wichtig, weil die großen Suchmaschinenhersteller Bing, Google, Yahoo! und Yandex dahinter stehen. Die Bedeutung wird eingeführt, indem unter `schema.org` auf standardisierte Weise Daten Begriffe zugeordnet werden, die für die Auszeichnung verwendet werden können. Die Syntax von `schema.org` wird von HTML Microdata<sup>2</sup> spezifiziert und ist dafür gedacht, maschinenlesbare Metainformationen in Webseiten zu integrieren. Die komplette Itemtype-Hierarchie von `schema.org` kann online abgerufen werden:

<http://schema.org/docs/full.html>

Ein Beispiel soll den Gebrauch von Microdata erläutern:

```
<div itemscope itemtype="http://schema.org/Book">
  <h3></span itemprop="genre">Lexikon</span><h3>
</div>
```

Das Element mit dem Attribut `itemscope` umgibt den kompletten Informationsblock. Stößt der Parser der Suchmaschine auf dieses Attribut, so weiß er, dass es sich um ein Microdata-Format handelt und erwartet dann eine weiterführende URL. Diese in `itemtype` enthaltene URL verweist nun auf den entsprechenden Eintrag unter `schema.org` und vermittelt der Suchmaschine die Bedeutung der enthaltenen Information. Die Eigenschaften des Buches werden dann durch das HTML-Attribut `itemprop` ausgezeichnet.

Für die Überprüfung von semantisch ausgezeichneten Webseiten gibt es ein Test-Tool für strukturierte Daten als Bestandteil der Google Webmaster-Tools (siehe Abschnitt 7).

<sup>2</sup> <http://www.whatwg.org/specs/web-apps/current-work/multipage/microdata.html#microdata>

## 6. Steuerung von Crawlern

### Datei robots.txt

- Webcrawler (Spider, Bots) sind Computerprogramme, die Websites regelmäßig besuchen und versuchen, deren Inhalte zu erfassen (indizieren). Normalerweise hat man nichts dagegen, weil man im Datenschwungel des Internet gefunden werden möchte, und versucht deshalb, den Crawlern die Arbeit möglichst zu erleichtern. Manchmal will man aber in Bezug auf Crawler selektiver vorgehen. Dazu dient die Datei robots.txt: In einer Datei dieses Namens können Betreiber von Websites angeben, welcher Such-Robot welche Inhalte (Verzeichnisse) auslesen darf und welcher was nicht lesen darf. Allerdings werden diese Anweisungen nur von seriösen Suchmaschinen beachtet. Alternativ kann das Crawling auch über ein entsprechendes Meta-Tag im Abschnitt <Head> einer Seite gesteuert werden (siehe Abschnitt 1). Da es keinen verbindlichen Standard für robots.txt gibt empfiehlt sich, die Syntax für die bevorzugte Suchmaschine in deren Dokumentation nachzuschlagen. Für Google findet man diese unter

[https://developers.google.com/webmasters/control-crawl-index/docs/robots\\_txt?hl=de](https://developers.google.com/webmasters/control-crawl-index/docs/robots_txt?hl=de)

Beispiel für robots.txt:

```
User-agent: *           | gilt für alle Webcrawler  
Disallow: /Temp/       | darf nicht gecrawled werden
```

Die Datei robots.txt kann hier getestet werden:

<https://support.google.com/webmasters/answer/156449?rd=1>

Weitergehende Erläuterungen findet man beispielsweise hier:

<http://www.seo-book.de/onpage/was-der-seo-uber-die-robots-txt-wissen-sollte>

oder hier:

<http://www.suchmaschinen-doktor.de/optimierung/meta-tags.html>

## 7. Google-Webmaster-Tools einsetzen

- In Deutschland ist Google die mit großem Abstand am meisten verwendete Suchmaschine (Marktanteil in 2013: 91%). Auf der Homepage des RatSWD wird die Google-Custom-Search eingerichtet. Es empfiehlt sich also, Websites für Google zu optimieren und dazu die Google-Webmaster-Tools einzusetzen. Das sind Analysen und Werkzeuge, um eine stabile und Google-freundliche Website erstellen zu können.

Sie finden die Webmaster-Tools unter:

<https://www.google.com/webmasters/>

## 8. Forschungsdaten

Die folgenden Hinweise beruhen auf Erfahrungen, die das ZPID mit dem Einsatz der Psychologie-Suchmaschine PsychSpider bei einer Reihe von Datenzentren gesammelt hat. Eigentlich sollte ihre Beachtung für Informationsinfrastruktureinrichtungen bzw. wissenschaftliche Informationsanbieter selbstverständlich sein, aber die Praxis belegt, dass es hier Verbesserungsbedarf gibt.

### Metadaten verwenden

- **Autor, Titel.** Gerade für nicht intuitiv verständliche Forschungsdaten gilt, dass die Dokumente (HTML, PDF, DOC) treffende Metadaten enthalten sollten. Die Erfahrung zeigt aber, dass PDF-Dateien häufig keine, unzureichende oder gar irreführende Metadaten (Autor und Titel) aufweisen.
- **Keywords.** Es empfiehlt sich, für Schlagwörter (Keywords) auf kontrolliertes Vokabular zurückzugreifen und dabei auch redundant vorzugehen, wenn es disziplinabhängig unterschiedliche Vokabulare gibt (z.B. Sozialwissenschaften und Medizin). Eingescannte Dokumente immer als Text oder als Image mit hinterlegtem Text (OCR) einbinden.

Beispiel:

Die Suche nach *Burnout* ergibt in der Trefferanzeige u.a. den Titel

Dis09122001\_A.PDF

Tatsächlich lautet der Titel des PDF-Dokuments, das auf einem universitären Dokumentenserver abgelegt ist (Dissertation)

Beanspruchung und Belastung der Altenpflege

### Nicht-Textuelle und interaktive Elemente vermeiden

- Vermeidung von Cookies, Frames, Scripting (Flash, JavaScript, HTML5, etc.). Datensätze sollten gut strukturiert und maschinenlesbar (barrierefrei), d.h. als Text aufbereitet werden (Schlagwörter, Stichprobengröße, Erhebungszeitraum, Population, usw.). Nur so ist eine automatisierte Indexierung gewährleistet.

### Umstrukturierung vermeiden

- Häufige Umstrukturierungen von Webauftritten erschweren die Auffindbarkeit von Inhalten ganz erheblich. Bei der Umstrukturierung die Suchmaschinenperspektive im Hinterkopf behalten (Aufwand für den Neuaufbau des Index<sup>3</sup>, Inkonsistenzen, einen Plan für 301-Redirects aufstellen).

---

<sup>3</sup> Das Crawling-Intervall kann je nach Suchmaschine mehrere Wochen bis Monate dauern.

## **Systematisches Web-Design**

- Eine Website, die von einem Crawler indiziert werden soll, stellt sich diesem als Sammlung von hypermedial verknüpften Texten, Bildern und anderen Multimedia-Inhalten dar. Damit aus einer Sammlung ein gut bedienbares und verständliches Ganzes wird, sind Erkenntnisse aus Gebieten wie Hypermedia-Design, Web-Design und Interface-Design hilfreich (Garrett, 2010). Aufgabe ist es, Knoten (Inhalte) und Verknüpfungen (Links) in hypermedialen Dokumenten so zu strukturieren, dass ein intuitiver und einfacher Zugang zu den Inhalten ermöglicht wird.
- Bei Websites von Datenzentren sollte es selbstverständlich sein, dass zusammengehörige Teile von Datensätzen wie allgemeine Studienbeschreibung (Kontext), Codebücher und die Daten selbst stets untereinander verlinkt werden!

## **Die Perspektive des Nutzers einnehmen**

- Beim Aufbau der Webseiten sollte berücksichtigt werden, welche Fragestellung die potentiellen Nutzerinnen und Nutzer haben könnten. Was könnten ihre Forschungsfragen sein? Welche Suchbegriffe werden sie verwenden? Man sollte beim Design die Perspektive der Nutzer einnehmen.
- Ein reales Beispiel soll dies erläutern. In einer medizinischen Datenerhebung werden physiologische Kennwerte wie Körpergröße und Körpergewicht sowie pathologische Befunde erhoben. Eine Arbeitsgruppe interessiert sich dafür, ob ein Zusammenhang zwischen Körpergewicht und Burnout-Erkrankung existiert. Wenn die relevanten Informationen (diagnostiziertes Burnout, Körpergewicht) jedoch so abgelegt sind, dass sie von Suchmaschinen als getrennte Inhalte erfasst werden, stellt sich das für Nutzer von Suchmaschinen so dar, als wären die Befunde nicht miteinander erhoben worden. Eine gute Webseitenstruktur, gute Metadaten und eine gute Verlinkung hätten diesem Irrtum entgegengewirkt.

## 9. Literatur

- Beel, J., Gipp, B., and Wilde, E. (2010). Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar and Co. *Journal of Scholarly Publishing*, 41 (2): 176–190. doi: 10.3138/jsp.41.2.176. University of Toronto Press. Downloaded from [www.docear.org](http://www.docear.org)
- Bing (2013): Bing Webmaster Guidelines. Retrieved from <http://www.bing.com/webmaster/help/webmaster-guidelines-30fba23a>
- Garrett, J. J. (2010). *The Elements of User Experience: User-Centered Design for the Web and Beyond* (2nd Edition). Berkeley: New Riders.
- Gattani, A. (). Web 3.0 and Semantic Search. Retrieved from <http://blog.kosmix.com/?p=1210>, 2010.
- Google (2011): Einführung in Suchmaschinenoptimierung. Retrieved from [https://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/www.google.de/de/de/webmasters/docs/einfuehrung-in-suchmaschinenoptimierung.pdf](https://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.de/de/de/webmasters/docs/einfuehrung-in-suchmaschinenoptimierung.pdf)
- Google (2013): Webmaster-Tools, allgemeine Richtlinien. Retrieved from [https://support.google.com/webmasters/topic/2370419?hl=de&ref\\_topic=1724125](https://support.google.com/webmasters/topic/2370419?hl=de&ref_topic=1724125)
- Wissenschaftsrat (2012): Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020 (WR Drs. 2359-12). Retrieved from <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>
- Seitenreport (2013): 23 detaillierte SEO Tipps für mehr Besucher. Retrieved from <http://www.seitenreport.de/kb/-/23-seo-tipps-fuer-mehr-besucher.html>
- Suchmaschinentricks (2013): Suchmaschinenoptimierung. Retrieved from <http://www.suchmaschinentricks.de/optimierung>
- Yandex (2013): webmaster help. Retrieved from <http://help.yandex.com/webmaster/?id=1113737>