



# Suchstrategien gegen die digitale Informationsflut: Finden und gefunden werden

Prof. Dr. Andreas Henrich – Lehrstuhl Medieninformatik

# Warum über Suche nachdenken?

- Google beantwortet täglich ca. 3,5 Milliarden Anfragen

[<http://www.live-counter.com/google-suchen/>, Stand 19.6.2015]

- 94% der Internet-Haushalte nutzen das Web für Kaufentscheide.

[Worldsites GmbH, CH-6343 Rotkreuz, <http://www.worldsites-schweiz.ch/suchmaschinen-bedeutung.htm>]

- According to a McKinsey report, employees spend 1.8 hours every day [...] searching and gathering information.

[<http://utrconf.com/top-3-reasons-why-we-spend-so-much-time-searching-for-information/>, Beth Burgee, May 2013]

# Überblick



im Web

4



im Unternehmen

1 Suchen  
bzw. Finden

2 Web Search  
&  
Web Intelligence

Enterprise  
Search  
(Wissensmanagement)

Weitere Suchen:

- Desktopsuche
- Shopsuche
- Bildsuche
- Site-Search
- ...

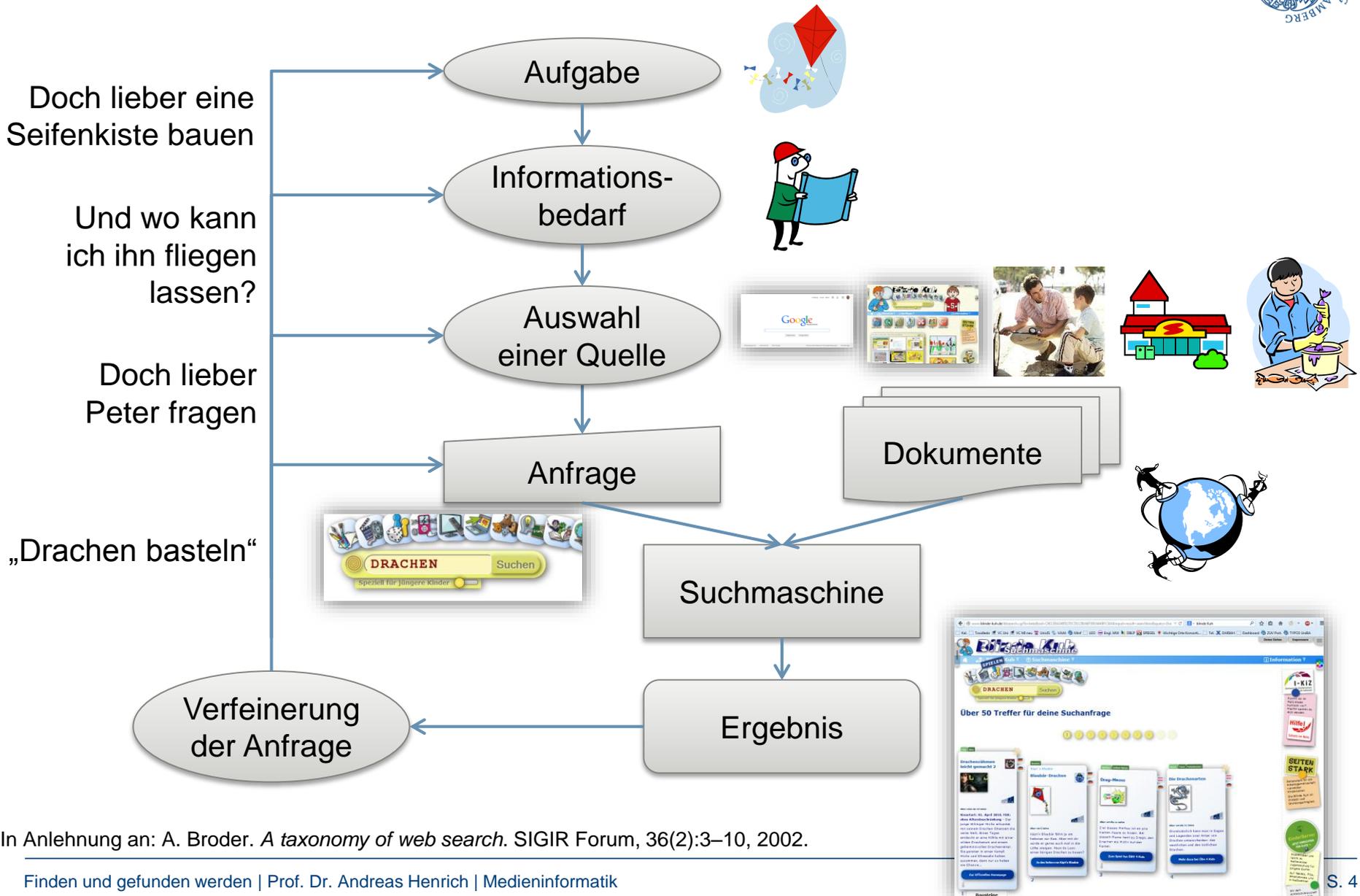
Gefunden  
werden

3 Search Engine  
Optimization

(E)CMS / KM

Bildquellen hier und im Weiteren wo nicht extra genannt: <https://openclipart.org/> sowie Microsoft PowerPoint ClipArts

# Was passiert bei einer Suche?



In Anlehnung an: A. Broder. *A taxonomy of web search*. SIGIR Forum, 36(2):3–10, 2002.

# Überblick

Die Situation:

- Milliarden von Dokumenten
- Relativ wenig Dateiformate
- Für viele / typische Anfragen große Zahl relevanter Dokumente
- starke Qualitätsunterschiede



# Typen von Suchtaktiken (Beispiele)

- **Monitor** – hilft fokussiert zu bleiben

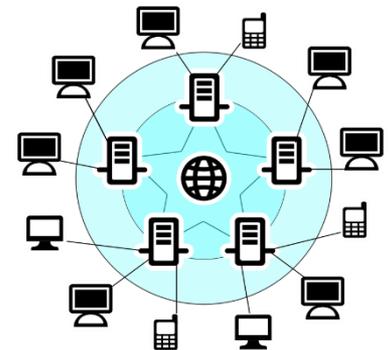
CHECK: Vergleich der ursprünglichen Suche mit der aktuellen

- File structure
- Formulation
- Term
- Evaluate



# Typen von Suchtaktiken (Beispiele)

- Monitor
- **File structure** – Nutzung der Struktur (des WWW)
  - BACKLINK: Suche Seiten, die eine „Perlenseite“ verlinken („link:...“)
  - TIMETRAVEL: Nutzung gecacheter Seiten (Wayback Machine)
- Formulation
- Term
- Evaluate

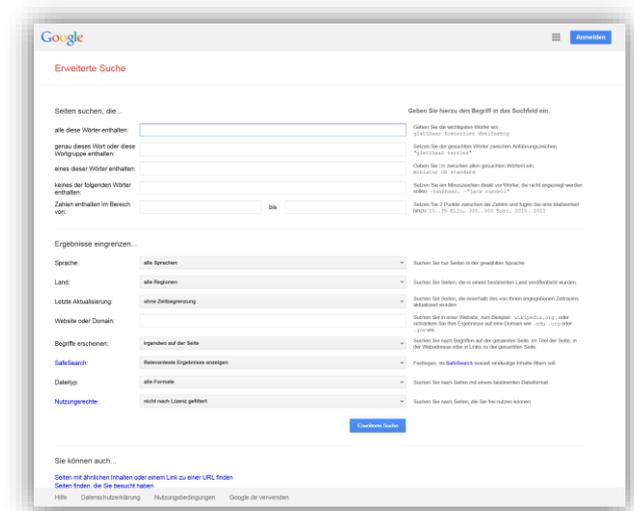


Nach: Alastair G. Smith, (2012) *Internet search tactics*, Online Information Review, Vol. 36 Iss: 1, pp.7-20

# Typen von Suchtaktiken (Beispiele)

- Monitor
- File structure
- **Formulation** – Nutzen der Funktionen der Suchmaschine  
BLOCK: **Ausschluss** gewisser Inhalte

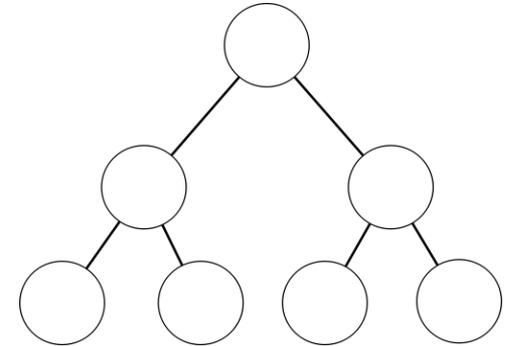
- Term
- Evaluate



Nach: Alastair G. Smith, (2012) *Internet search tactics*, Online Information Review, Vol. 36 Iss: 1, pp.7-20

# Typen von Suchtaktiken (Beispiele)

- Monitor
- File structure
- Formulation
- **Term** – Nutzung von Begriffsrelationen, Fachbegriffen
  - SUB: Verwendung **speziellerer Begriffe** (subordinate)
  - CONTRARY: Verwenden des logischen **Gegensatzes**
  - ANTICITERM: **Antizipieren** von Begriffen aus relevanten Seiten
- Evaluate



# Typen von Suchtaktiken (Beispiele)

- Monitor
- File structure
- Formulation
- Term
- **Evaluate** – Querbezüge herstellen, Ergebnisse hinterfragen  
CROSSCHECK: **Prüfen** der Informationen mit anderen Quellen



# Überblick

Die Situation:

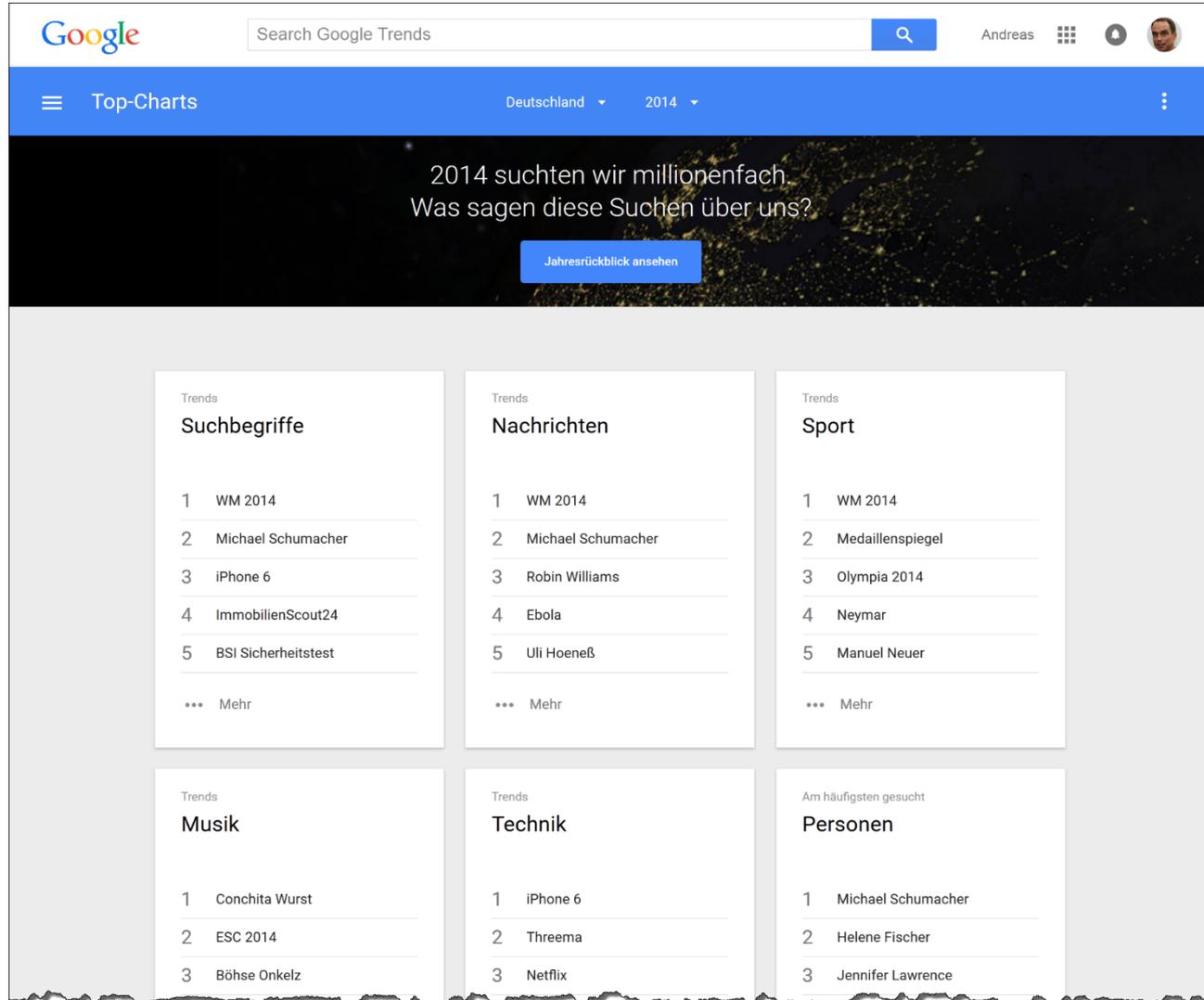
- Milliarden von Dokumenten
- Relativ wenig Dateiformate
- Für viele / typische Anfragen große Zahl relevanter Dokumente
- starke Qualitätsunterschiede



im Web



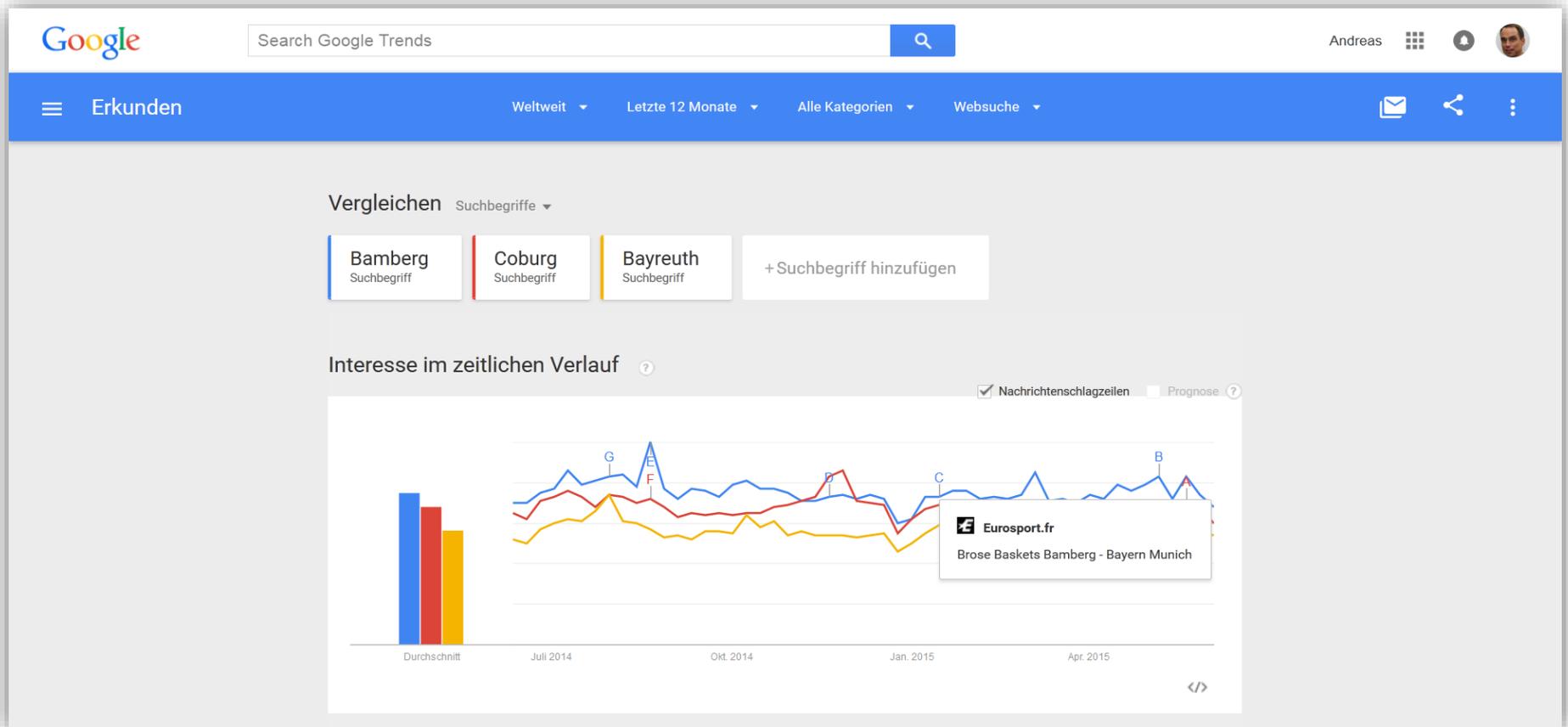
# Web Analytics



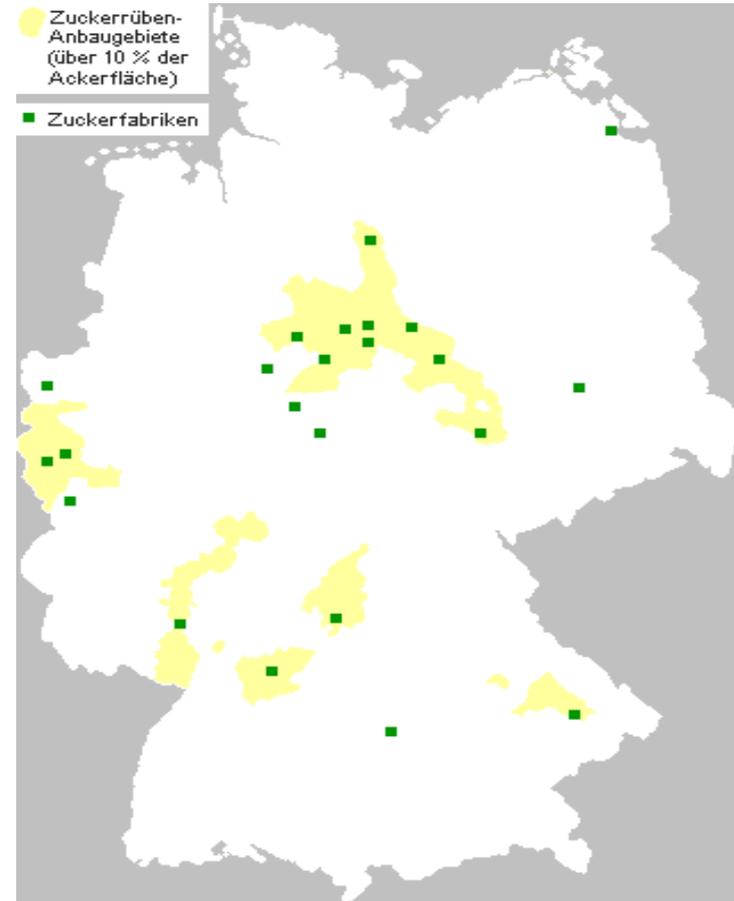
The screenshot displays the Google Trends interface for the year 2014 in Germany. The main heading reads "2014 suchten wir millionenfach. Was sagen diese Suchen über uns?" with a button to "Jahresrückblick ansehen". Below this, six trend categories are shown:

- Suchbegriffe**
  - 1 WM 2014
  - 2 Michael Schumacher
  - 3 iPhone 6
  - 4 ImmobilienScout24
  - 5 BSI Sicherheitstest
- Nachrichten**
  - 1 WM 2014
  - 2 Michael Schumacher
  - 3 Robin Williams
  - 4 Ebola
  - 5 Uli Hoeneß
- Sport**
  - 1 WM 2014
  - 2 Medaillenspiegel
  - 3 Olympia 2014
  - 4 Neymar
  - 5 Manuel Neuer
- Musik**
  - 1 Conchita Wurst
  - 2 ESC 2014
  - 3 Böhse Onkelz
- Technik**
  - 1 iPhone 6
  - 2 Threema
  - 3 Netflix
- Am häufigsten gesucht Personen**
  - 1 Michael Schumacher
  - 2 Helene Fischer
  - 3 Jennifer Lawrence

# Google Trends



# Beispiel: „Zuckerrübenanbau“



Bildquelle: <https://de.wikipedia.org/wiki/Zuckerr%C3%BCbe#/media/File:Zuckerr%C3%BCbengebiete-D.PNG>

Datenquelle: Quelle: Wirtschaftliche Vereinigung Zucker - Verein der Zuckerindustrie

# Überblick



im Web

Die Situation:

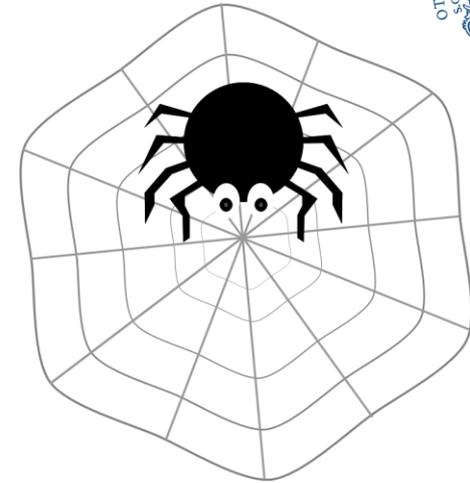
- Milliarden von Dokumenten
- Relativ wenig Dateiformate
- Für viele / typische Anfragen große Zahl relevanter Dokumente
- starke Qualitätsunterschiede

Web Search  
&  
Web Intelligence

Search Engine  
Optimization

# Search Engine Optimization

## Google Tipps 1/3



- Make your site **crawlable** by engines
- Allow search engines to crawl in **robots.txt**
- Each page can be reached by clicking **links**
- Use **standard HTML**: less Flash, AJAX, ActiceX

Quelle: Matt Cutts, Google, SEO-Search Inside, 13.01.2014, [https://youtu.be/LQI\\_0L\\_ggno](https://youtu.be/LQI_0L_ggno)

# Search Engine Optimization

## Google Tipps 2/3



- Pages should have clear **titles** and **descriptions**
- Use **text**, not just images of text
- Think about **what users will type**
- Example – **Search terms** for:



Quelle: Matt Cutts, Google, SEO-Search Inside, 13.01.2014, [https://youtu.be/LQI\\_0L\\_ggno](https://youtu.be/LQI_0L_ggno)

# Promoting your site

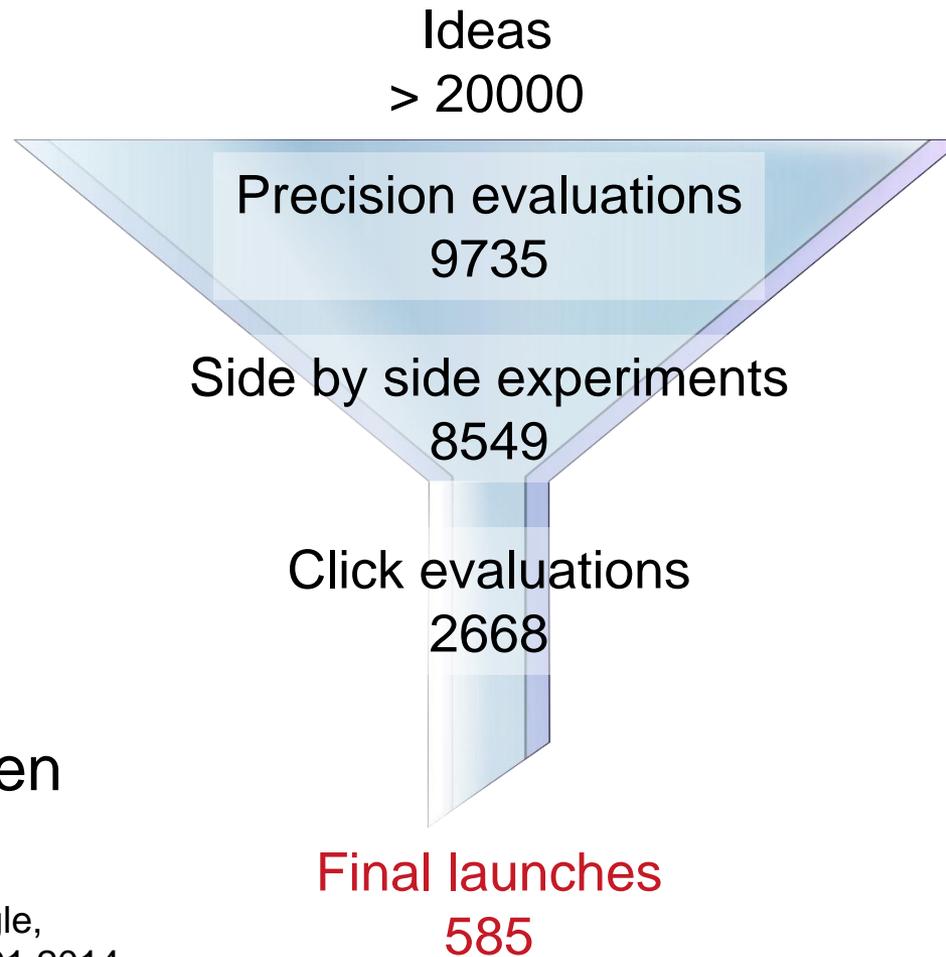
## Google Tipps 3/3



- Links matter, but don't spam to get links
- Good ways to get links: compelling content, blogging, newspaper stories.
- Social media can amplify your message.
- Free tools (z.B. <https://www.google.com/webmasters/>)

Quelle: Matt Cutts, Google, SEO-Search Inside, 13.01.2014, [https://youtu.be/LQI\\_0L\\_ggno](https://youtu.be/LQI_0L_ggno)

# Wie oft ändert Google seine Kriterien?



## Jährliche Zahlen (2011)

Quelle: Matt Cutts, Google,  
SEO-Search Inside, 13.01.2014,  
[https://youtu.be/LQI\\_0L\\_ggno](https://youtu.be/LQI_0L_ggno)

# On-The-Page SEO

# Off-The-Page SEO

On-The-Page SEO			Off-The-Page SEO			
Content	Architecture	HTML	Trust	Links	Personal	Social
<b>Cq</b> +3 Quality	<b>Ac</b> +3 Crawl	<b>Ht</b> +3 Titles	<b>Ta</b> +3 Authority	<b>Lq</b> +3 Quality	<b>Pc</b> +3 Country	<b>Sr</b> +2 Reputation
<b>Cr</b> +3 Research	<b>Ad</b> +2 Duplicate	<b>Hd</b> +2 Description	<b>Te</b> +2 Engage	<b>Lt</b> +2 Text	<b>Pl</b> +3 Locality	<b>Ss</b> +1 Shares
<b>Cw</b> +2 Words	<b>Am</b> +2 Mobile	<b>Hs</b> +2 Structure	<b>Th</b> +2 History	<b>Ln</b> +1 Numbers	<b>Ph</b> +3 History	
<b>Cf</b> +2 Fresh	<b>As</b> +1 Speed	<b>Hh</b> +1 Headers	<b>Ti</b> +1 Identity	<b>Vp</b> -3 Paid	<b>Ps</b> +2 Social	
<b>Cv</b> +2 Vertical	<b>Au</b> +1 URLs	<b>Vs</b> -2 Stuffing	<b>Vd</b> -1 Privacy	<b>VI</b> -3 Spam		
<b>Ca</b> +1 Answers	<b>Ah</b> +1 HTTPS	<b>Vh</b> -1 Hidden	<b>Va</b> -1 Ads			
<b>Vt</b> -2 Thin	<b>Vc</b> -3 Cloaking					

<http://searchengineland.com/guide/seo>

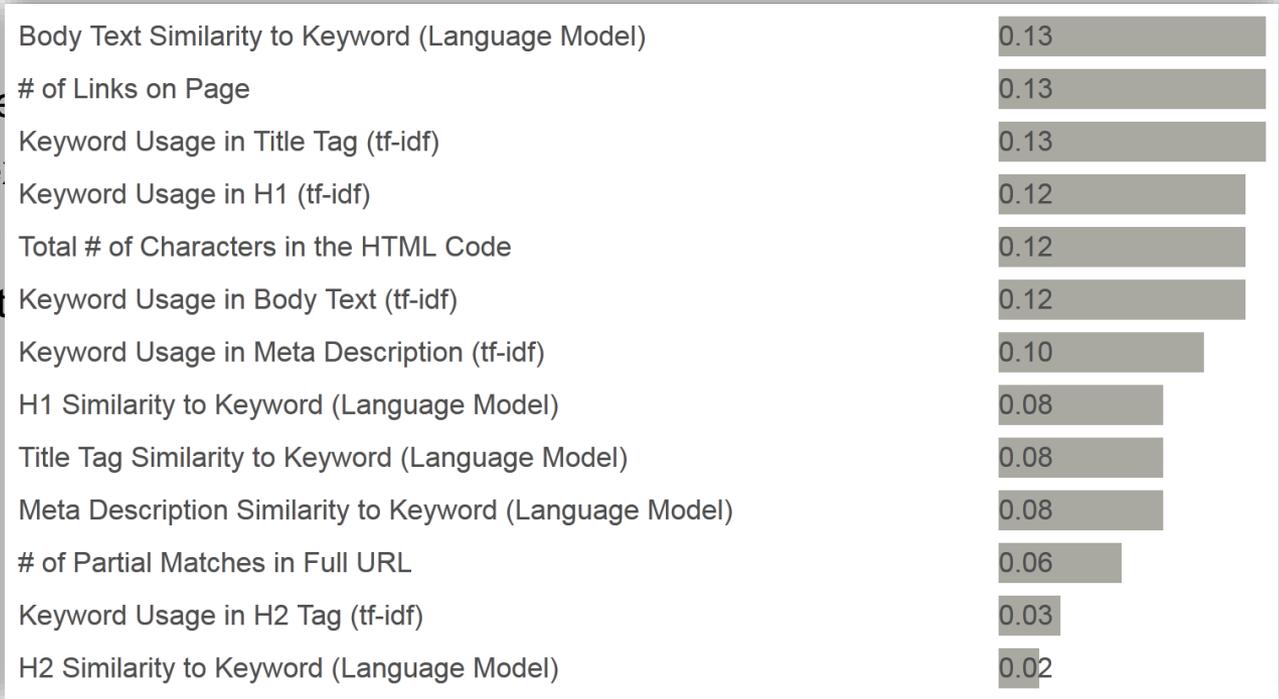
# SEO: Important Signals (Ranking Factors)



- Page Link Authority Features
- Page Level Social
- Page Level Anchor Text
- Page Keyword Usage
- Page Keyword Agnostic
- Domain Link Authority Features
- Domain Level Anchor Text
- Domain Keyword Usage
- Domain Keyword Agnostic
- Brand Metrics

# SEO: Important Signals (Ranking Factors)

- Page Link Authority Features
- Page Level Social
- Page Level Anchor Text
- ▶ Page Keyword Usage
- Page Keyword Agnostic
- Domain Link Authority Fe
- Domain Level Anchor Te
- Domain Keyword Usage
- Domain Keyword Agnost
- Brand Metrics

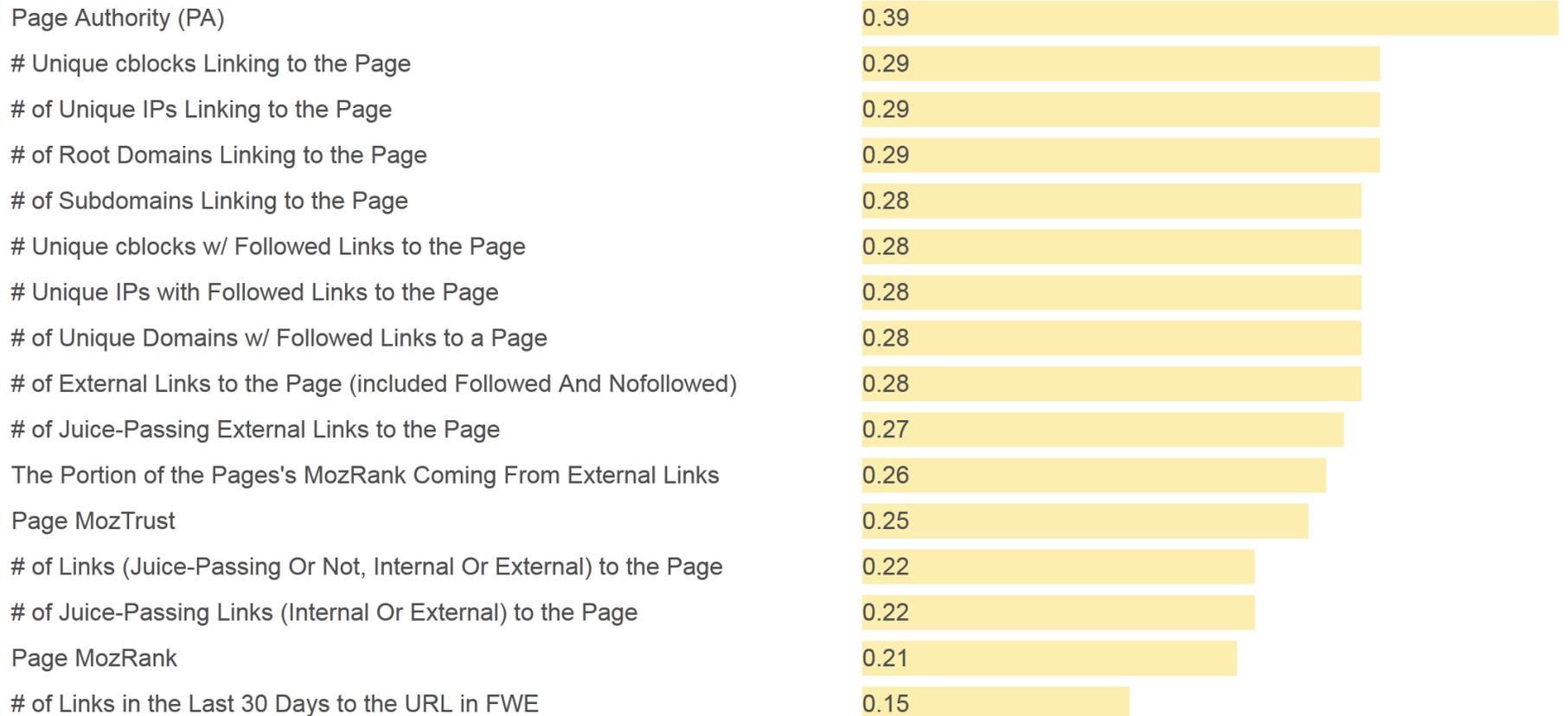


# SEO: Important Signals (Ranking Factors)



## ► Page Link Authority Features

### • Page Level Social



# Überblick

Die Situation:

- Viele verschiedene **Dateiformate**
- Viele verschiedene betroffene **Systeme**
- **Zugriffsrechte / Vertraulichkeit** (auch der Suche)
- **Keine** ausnutzbare **Linkstruktur**
- i.d.R. **wenig relevante Dokumente**
- **keine SEO-Incentivierung** der Autoren
- **Architektur: Gathering / Metasearch / ...**



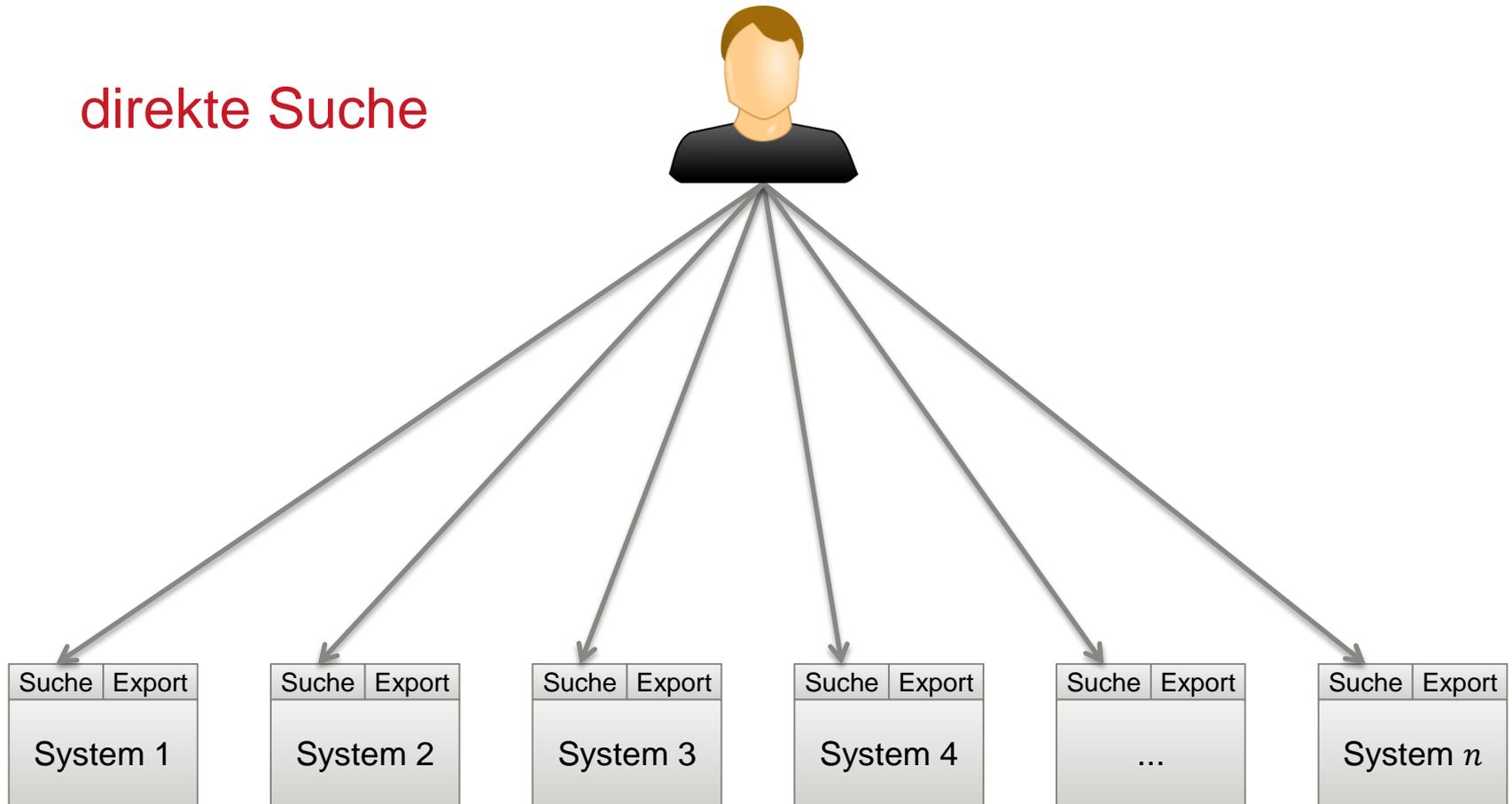
im Unternehmen

Enterprise  
Search  
(Wissensmanagement)

(E)CMS / KM

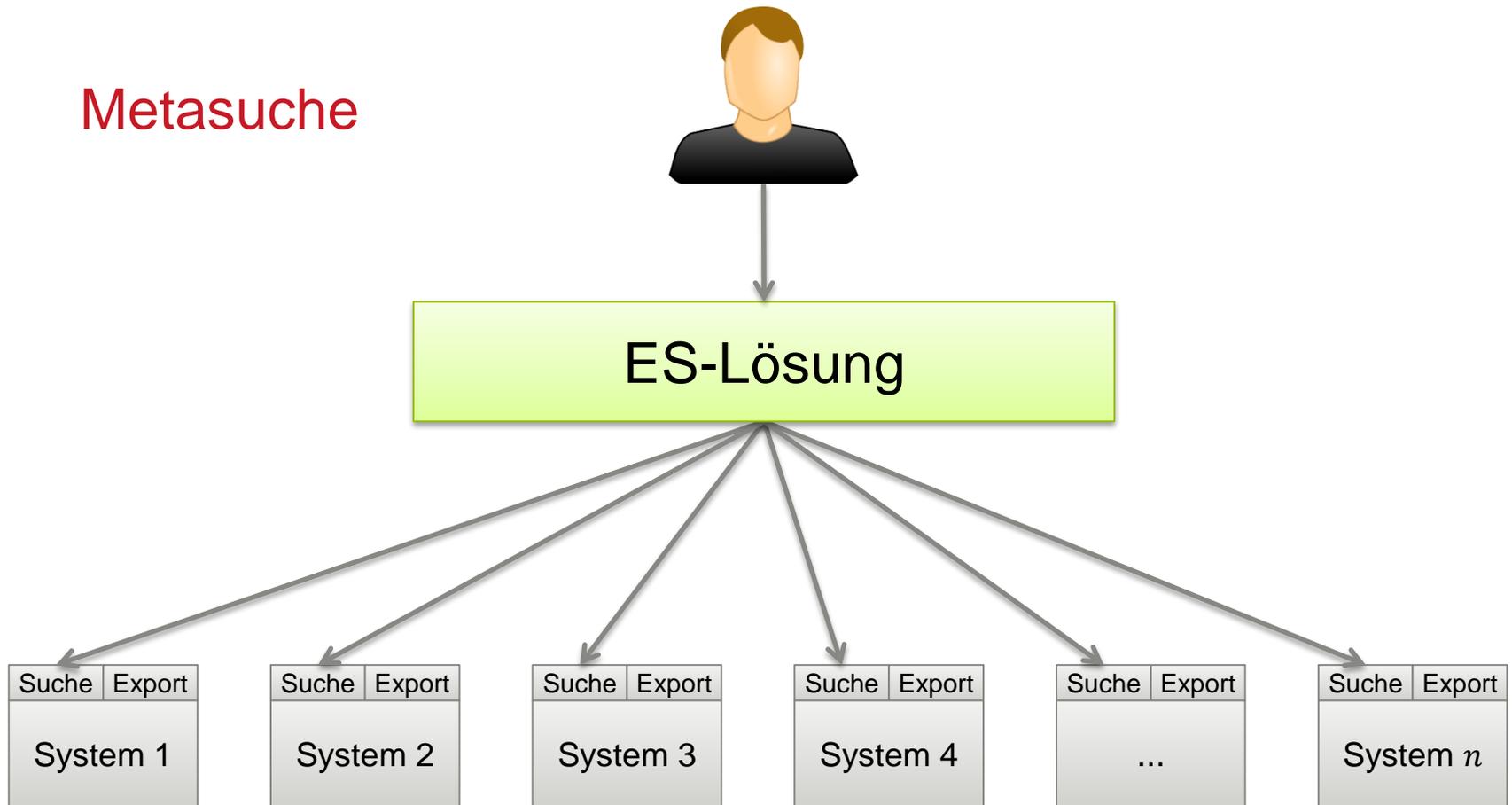
# Alternativen

direkte Suche



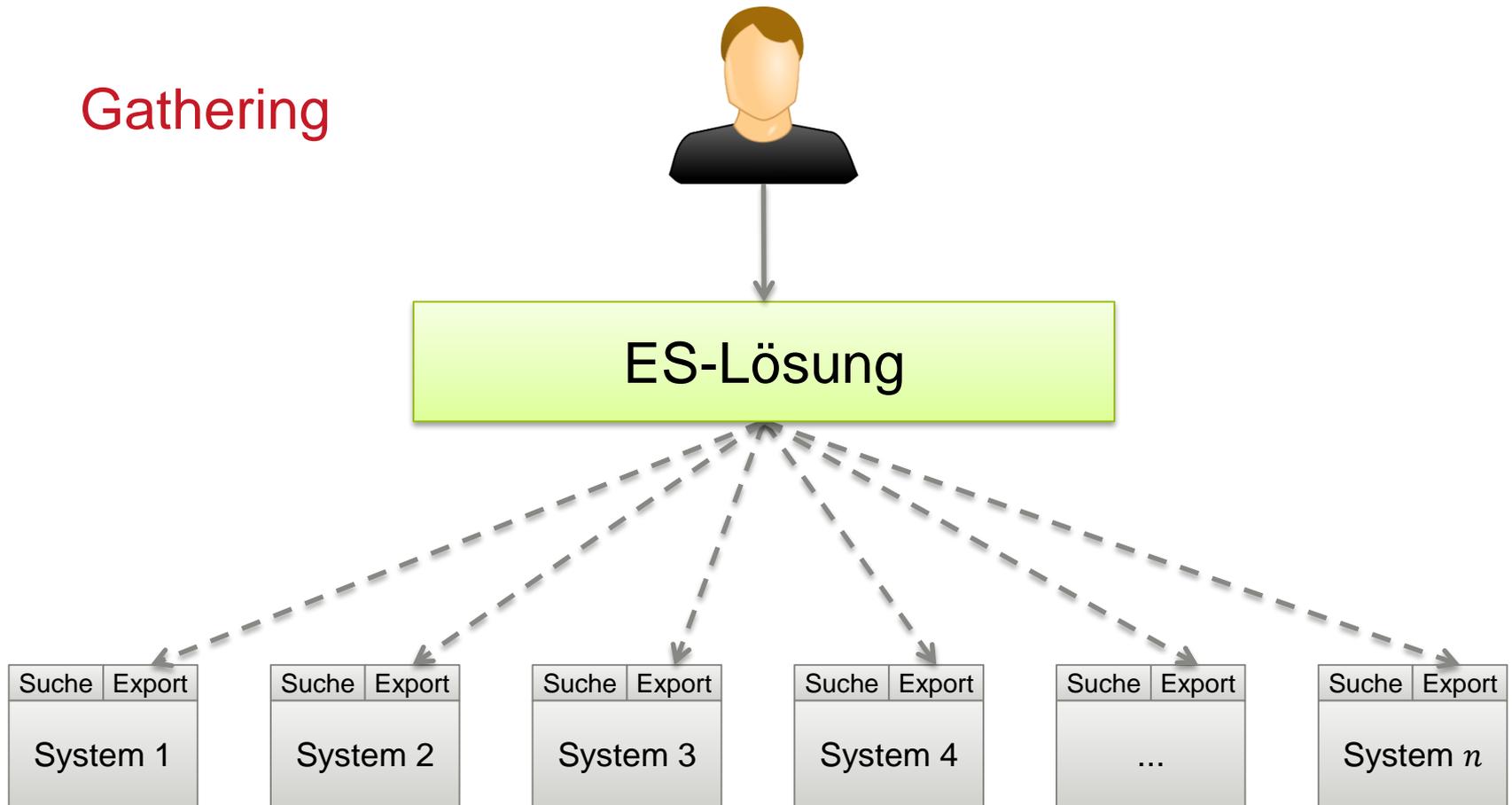
# Alternativen

## Metasuche



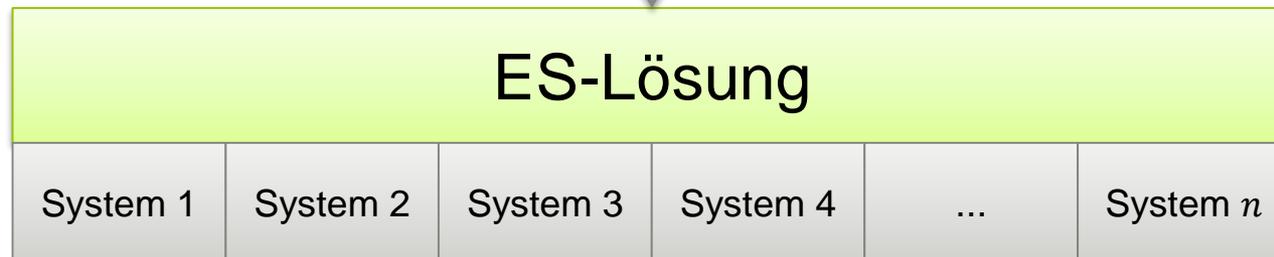
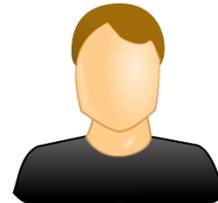
# Alternativen

## Gathering

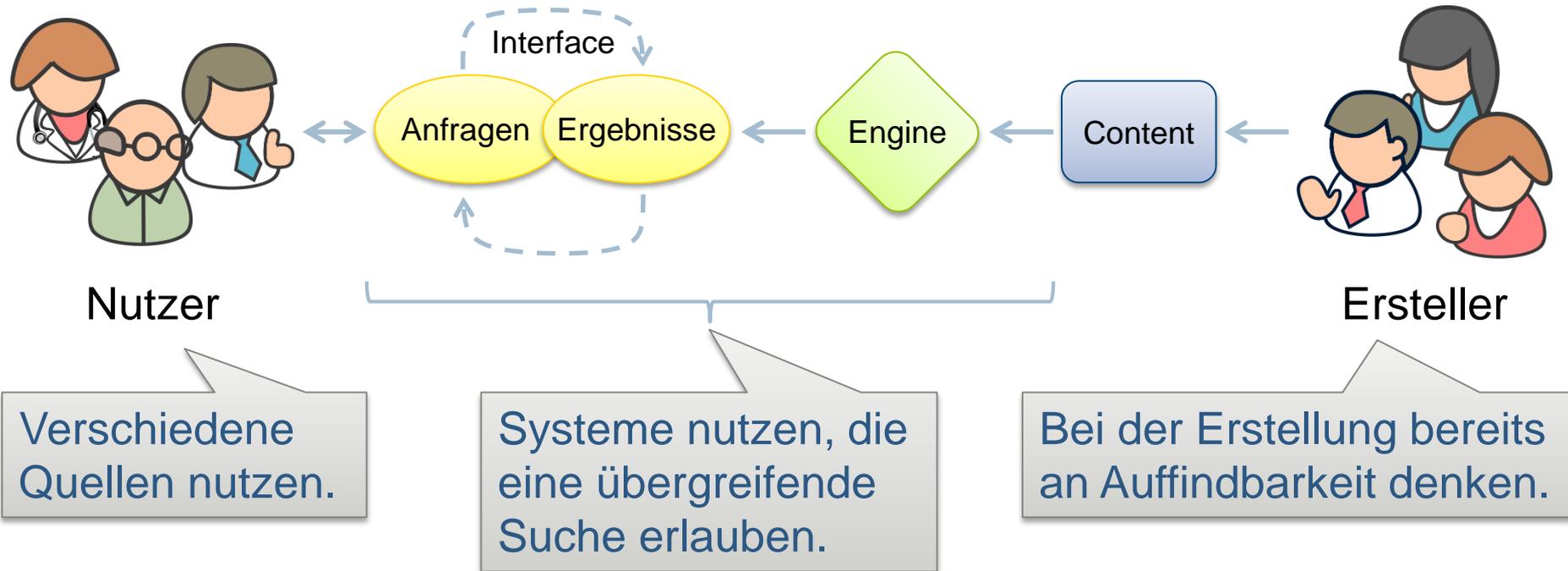


# Alternativen

integrierte Lösung



# Man muss an vielen Stellen ansetzen!



In Anlehnung an: Morville, Peter; Callender, Jeffery (2010): Search patterns. Sebastopol, CA: O'Reilly.

# Social Media kann Enterprise Search helfen

## Social Media Ansätze in

- Intranet
- Web-Portal
- Hilfeforen

## Ziele:

- vorhandene Inhalte anreichern
  - **Bewertung** eines Dokuments
  - **Verschlagwortung** durch die Nutzer (Tagging)
  - **kollaborative Erstellung** (von Dokumenten / Datenbasen)



Nach: Lewandowski, Dirk (2010): Whitepaper Enterprise Search: Was die Nutzer erwarten und warum Social Media so entscheidend ist. Hg. v. T-Systems Multimedia Solutions GmbH. Dresden.

# Zusammenfassung

