

Gutenberg 3.2 – Ebook-Piracy Report

What does piracy cost the publishers? – „Ersatzraten“ (replacement rates)

Manuel Bonik

Dr. Andreas Schaale



Pic. from ref. [1]

Berlin, September 2012

Prologue

The report Gutenberg 3.2 presents a continuation of the versions 3.0 and 3.1 [2-4] published mainly for the German book and ebook market.

In 2011 and 2012 Internet piracy is an ongoing issue for the content industry (music, movies, books ...) as well as for the international policy and last but not least for the Internet users. We will demonstrate how the illegal distribution of ebooks has developed in the last 1-2 years. Our estimations and conclusions are based on empirical data, derived from our daily business. It is (for us) not possible to reveal every aspect of the problem in general, but we can and will show some examples, how the (illegal) distribution and consumption of ebooks are evolving in this time.

General Trends

In the report [3] we have given number of the traffic (usage) of one of the largest German sites that distributes ebooks together with other content formats. The following screenshot shows the views of one thread of the community www.b***.bz which offers many thousands ebooks in German language. The ebook files are not stored at the servers of the community but on so called one click hoster (OCH) which offer file storage for free. Most of the links of the example below lead to the OCH uploaded.to.

Screenshot of b***.bz (threads with ebook offers), Jan. 2011



Thema / Autor	Letzter Beitrag	Beiträge	Hits	Forum
10 000 eBooks und mehr - update vom 07.12.10 (1 2 3 4) Dana Frost, 06.10.08	Heute 22:28 von Gilhalad	88	1.049.950	Magazine und Literatur-Dauerangebote
Gesamtübersicht Digitale Bibliothek - update vom 07.12.10! (1 2 3) Lüdegast, 08.10.08	Heute 21:04 von Kalopsi	62	371.862	Magazine und Literatur-Dauerangebote
Spiegelbestseller Belletristik (Hardcover+ Taschenbuch) [02/2011] spiegelbest, 30.04.10	Heute 11:33 von Quacko1	5	161.160	Magazine und Literatur-Dauerangebote

The main thread on the top has had at that time about 1 million views (from Oct. 2008 until Jan. 2011). This number can be compared with the same thread about 1.5 years later:

Screenshot of b***.bz (threads with ebook offers), Sep. 2012



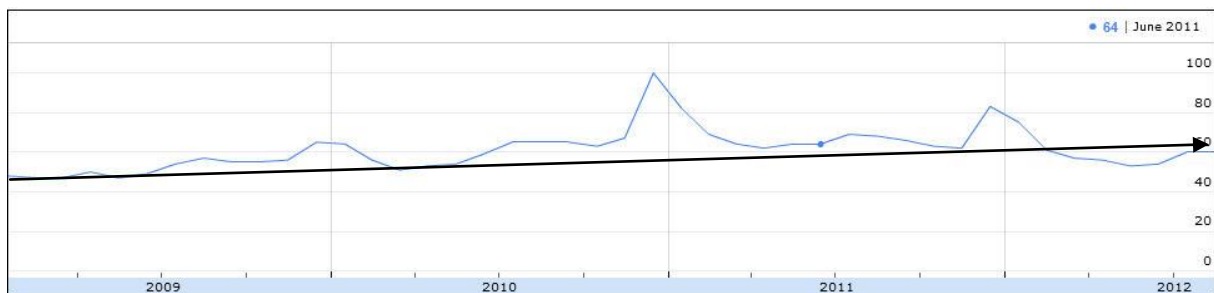
Thema / Autor	Letzter Beitrag	Beiträge	Hits	Forum
10 000 eBooks und mehr - update vom 31.05.12 (1 2 3 4 5 ... Letzte Seite) Dana Frost, 06.10.08	Heute 22:17 von Lüdegast	140	3.201.388	Magazine und Literatur-Dauerangebote

In the last 1.5 years this thread has about 2.2 million new hits. The speed of growth has increased dramatically, compared to 2011. This demonstrates the growing interest of the users to get ebooks free of charge. It is worth to be mentioned, that the whole community [b***.bz](http://www.b***.bz) now has more than 2 million registered (German speaking) users. It is among the top 100 (pos. 91, worldwide pos. 1.500) leading sites in Germany (by traffic rank [5]). One can estimate the amount of ebooks distributed by

this single thread: The ebooks are organized in packages of approximately 15 ebooks per single OCH download link. The 2.2 million hits in the last 1.5 years correspond to about 10-30 million ebooks distributed in the time of 1.5 years. This number exceeds the amount of ebooks which have been sold in Germany by a multiple. In 2011 about 4.7 million ebooks have been sold in Germany [6]. The fact, that a single thread of one community distributes at least twice as much ebooks illegally than the German ebook market overall, shows the malady of the regular market in Germany.

The problem of ebook piracy is not a problem of Germany only. The search for ebooks in the world is increasing.

Google trends of searches for “ebooks” from Jan. 2009 until Aug. 2012 (Germany)



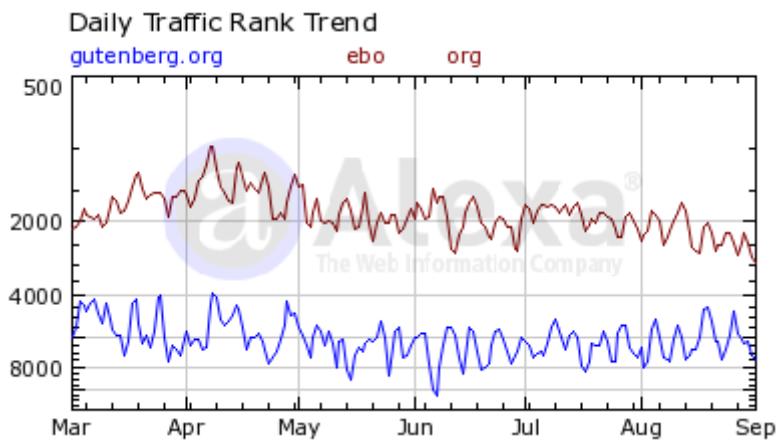
Besides the growth of the search for certain keywords (ebooks) Google also offers the information of the search terms, the context, where these keywords appear.

Related keywords for “ebooks”, Google trends (Germany)

Search terms	
Top searches	
1. ebooks free	100
2. download ebooks	55
3. ebook	35
4. books	30
5. pdf ebooks	20
6. ebooks for free	20
7. kindle ebooks	20
8. kindle	20
9. free books	20
10. free ebook	15

One can clearly recognize that most of the search terms do not express the intention to *buy* ebooks! Of course, there are legal offers of free ebooks. The Projekt Gutenberg [7] is the world’s largest collection of ebooks, which copyright has expired. Comparing the traffic rank of the (legal) Projekt Gutenberg with one major ebook “warez” site offering illegal downloads, one will get the following picture:

Alexa traffic rank of Gutenberg.org vs. ebo***.org



One can easily see, that the traffic on the illegal (one of many) site ebo***.org is significantly larger than the traffic on Gutenberg.org. The users searching for free ebooks are mainly searching for (illegal) copies of modern ebooks.

An interesting question is, how many ebooks (with copyrights) are available and distributed via illegal channels? Of course, nobody can answer this question exactly – the Internet is too big and very dynamic. However we could perform a test on a Russian based “library” which collects ebook copies and exchanges them with other (also illegal) sites. This activity is called “Library Genesis” and it operates with different URL’s. Downloading their database officially (just the metadata, not the actual ebooks) we could find the following picture (status June 2012):

Top 20 list of the Library Genesis

Publisher or Imprint	Nb. of books
Springer	55.623
Wiley	22.745
Cambridge	14.376
Oxford University Press	10.696
Routledge	9.095
CRC	7.243
Mcgraw Hill	7.233
Academic Press	5.482
Palgrave MacMillan	4.802
Elsevier	4.244
World Scientific	3.520
Osprey	3.236
Prentice Hall	3.066
O'Reilly	3.047
Humana	2.825
MIT Press	2.770
Addison Wesley	2.553
Penguin	2.544
Brill	2.308
Princeton University Press	2.238

All together this site offers 824.000 ebooks (June 2012) in more than 10 languages. Most of the books belong to well known STM publishers. A recent test has shown, that a site related to the Genesis library (lib***.info or book***.org) now offers more than a million ebooks including fiction and non fiction:

Screenshot of one distribution site related to the Library Genesis



Of course, those numbers of single sites are not representative for the whole Internet, but they determine a *lower limit* of available ebooks. We estimate that at present the amount of illegal ebook titles in the Internet is in the region of 3-6 millions. Of course, each title is to be found usually in several blogs with a few independent mirrors per title or edition.

Finally one can assume that a few million books are available on typical “piracy sites.” Usually the communities where those books are offered do not host the books, but they offer links to OCH’s (today there exist at least 300 of them, with new ones coming up every week) where the interested users can perform the actual download. (In the ebook piracy so called P2P exchange systems (like bittorrent, edonkey, ...) play a minor role. Due to our observations only a tiny fraction (less than 5%) of ebook copies are distributed via these systems.) The number of users downloading those illegal book copies is growing, correlated to the sales of reading devices such as Kindle, iPad etc. Estimates of iPad sales indicate that until now about 20 million devices have been sold [8]. Keeping in mind, that there exist similar reading devices (Kindle, ...) the amount of users which own an ebook reader will reach the number of 100 million soon. Among students the usage of that kind of ebooks is spreading, because textbooks are expensive and many students own notebooks. It’s no secret, that many of those illegally distributed textbooks are via “Copy&Paste” the source for producing homeworks or even diploma and PhD thesis. Unlike official ebooks usually with a DRM protection, illegal ebooks make it easy to produce copies of the texts. Together with the “low” price, this makes those ebooks attractive for students. Reading the posts in certain (not only academic) communities it seems, that getting university textbooks this way is very common practice, saving time and money. In certain cases even the professors recommend students to get their material from those piracy sites for free.

All of this shows that the piracy of books continues growing further. It’s only a question of time, when the regular sales will become some kind of “rare exception.” Not only for STM publishers!

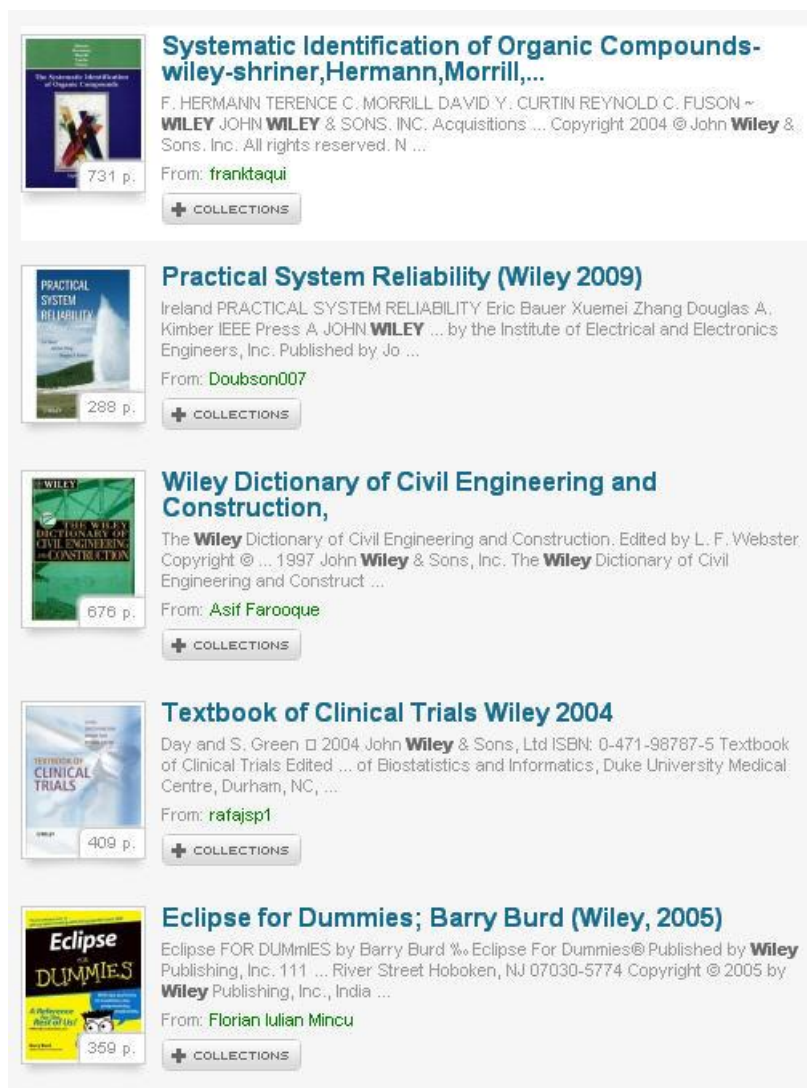
New forms of ebook piracy

In the report above we have studied aspects of “regular” Internet piracy only. Regular piracy means, that the communities and their users are anonymous and that the files are not stored on the servers of the communities, but on OCH’s. People visiting those communities, uploading and downloading ebooks are mostly aware of the fact, that this activity violates the copyrights (at least the uploads). Besides this “regular piracy” there exist forms of distributing copyrighted material which are not obviously related to piracy activities – however, they are.

Scribd.com

Scribd is a regular company which is located in the US. They claim to be “the world’s largest online library.” Well, with an Alexa traffic rank of about 200 they might indeed be the “YouTube for documents.” Besides completely legal material, like presentations, articles or other texts, Scribd offers copies of complete books which are definitely a violation of the copyrights. Our example is Wiley:

Screenshot of "free" Wiley books at Scribd.com



The screenshot displays a list of five books available for free on Scribd.com. Each entry includes a book cover, the title, author information, a page count, and a 'COLLECTIONS' button.

- Systematic Identification of Organic Compounds-wiley-shriner,Hermann,Morrill,...**
F. HERMANN TERENCE C. MORRILL DAVID Y. CURTIN REYNOLD C. FUSON ~ WILEY JOHN WILEY & SONS, INC. Acquisitions ... Copyright 2004 © John Wiley & Sons, Inc. All rights reserved. N ...
From: [franktaqui](#)
731 p.
+ COLLECTIONS
- Practical System Reliability (Wiley 2009)**
Ireland PRACTICAL SYSTEM RELIABILITY Eric Bauer Xuemei Zhang Douglas A. Kimber IEEE Press A JOHN WILEY ... by the Institute of Electrical and Electronics Engineers, Inc. Published by Jo ...
From: [Doubson007](#)
288 p.
+ COLLECTIONS
- Wiley Dictionary of Civil Engineering and Construction,**
The **Wiley** Dictionary of Civil Engineering and Construction. Edited by L. F. Webster. Copyright © ... 1997 John **Wiley** & Sons, Inc. The **Wiley** Dictionary of Civil Engineering and Construct ...
From: [Asif Farooque](#)
676 p.
+ COLLECTIONS
- Textbook of Clinical Trials Wiley 2004**
Day and S. Green © 2004 John **Wiley** & Sons, Ltd ISBN: 0-471-98787-5 Textbook of Clinical Trials Edited ... of Biostatistics and Informatics, Duke University Medical Centre, Durham, NC, ...
From: [rafajsp1](#)
409 p.
+ COLLECTIONS
- Eclipse for Dummies; Barry Burd (Wiley, 2005)**
Eclipse FOR DUMMIES by Barry Burd ® Eclipse For Dummies® Published by **Wiley** Publishing, Inc. 111 ... River Street Hoboken, NJ 07030-5774 Copyright © 2005 by **Wiley** Publishing, Inc., India ...
From: [Florian Iulian Mincu](#)
359 p.
+ COLLECTIONS

As far as Wiley is officially selling ebooks on Scribd (in the paid section), we have reasons to believe, that those copies (there are more than 1000 of them) are not authorized by the publisher. Wiley is just one example. On Scribd you will find complete copies of books of the STM book publishers and fiction books as well. It is worth to be mentioned, that Scribd allows to read books only when you are online on their site. Downloading those books requires a paid account. However, there are numerous hacks, how to rip (download and save) the documents which are "read only" for free. One can question if the holders of Scribd are aware that their site is one of the world's biggest piracy sites at the moment. In any case it shows the limits of legal sales of ebooks in the Internet. Pandora's box is open.

“Flat rate shops”

Piracy appears in different forms. One (rather new) form is the so called “flat rate shop.” Here is an example:

[Example of a flat rate shop - online-library.ws](#)

BOOKS ONLINE

HOME EBOOKS AUDIO BOOKS LOG IN ACCOUNT CONTACT

HOME LOG IN WROX

SEARCH

Search: wrox eBooks **SEARCH**

EBOOKS

Professional SQL Server 2000 Database Design
Louis Davidson, Wrox Author Team

<p>Database design is one of the most contentious issues in computer science. There is always a delicate balance to be struc...[read more](#)

eBook

Beginning ASP.NET Security (Wrox Programmer to Programmer)
Barry Dorrans

Beginning ASP.NET Security (Wrox Programmer to Programmer)...[read more](#)

eBook

Publication Manual of the American Psycholog...
American Psychological Association

The 7 Habits of Highly Effective People
Stephen R. Covey

The Official Guide for GMAT Review
Graduate Management Admission Coun...

The China Study: The Most Comprehensive Stud...
Howard Lyman, John Robbins, T. Col...

How to Read Literature Like a Professor. A L...
Thomas C. Foster

Too Big to Fail: The Inside Story of How Wal...

Paying a subscription fee of 39,90\$ a month the user gets access to 200.000 ebooks and 20.000 audio books and may download them – it’s a kind of “All you can read” offer. Of course, these books are not licensed by the copyright holders! At present we have observed more than 100 sites selling subscriptions to this kind of “library service.”

This new method of piracy works without advertising payments. Following the discussions of illegal ebook distributions (at amazon, ebay etc.) we are astonished, that none of the content owners has detected that kind of business model, yet. For users (academic or non academic) this flat fee is very convenient, because these shops are free of advertisement and there are no download restrictions (most of the OCH’s do not allow unlimited free downloads for free; of course there are tools like the highly popular JDownloader to work around these limitations). We assume that most of the users of this kind of “shops” are not aware of the fact, that these sites are illegal. This way the flat rate shops might attract people, which in general would avoid using typical “warez” sites.

Ersatzraten – Replacement Rates

“There is no such thing as a free lunch.” Obviously there is! If you can get an expensive book for free – why should you buy it? The “free book lunch” costs the publishers (and authors and bookshops and ...) money. Sales are replaced by the consumption of free ebooks – Ersatz!

There are only little doubts, that the availability of free alternatives will cause a decrease of sales. This might not be true for all the authors and all their books – maybe sometimes it is helpful to distribute books for free as a marketing or promotion tool. If we are speaking about well-known authors or textbooks required for your academic work, it’s naive to believe that illegal copies will increase the sales of an ebook. The music industry has lost half of their revenues since 2000 – because of the piracy! Film industry also knows the problem well. The book publishers will follow soon. Statistical data already indicate that trend. In Germany the sales of the Top 3 best sellers dropped in 2011 by 27% (compared to 2010) and the revenues of the top 30 have dropped by 29% [9]. This is one example of lost sales.

In the following we will give an estimate, how much that might cost.

As the base for our estimates we use the download data provided by scribd.com, the largest site (by traffic) where one can view and download complete ebooks. As an example we again chose Wiley – this problem is by far not limited to this particular publisher. Besides the title, Scribd offers the following information: number of views and the date of the upload. Here is an example of the top “sellers” of Wiley at scribd.com:

Selected Wiley books at scribd.com on Aug. 4, 2012

Title	Hits	Date	Today	Days	Hits per day
John.wiley.and.Sons.umts.Performance.measurement.sep	25.984	27.10.2011	04.08.2012	282	92,1
Machowski-Bialek-Bumbi - Power System Dynamics - Stability and Control	21.104	27.02.2011	04.08.2012	524	40,3
Wiley .Verilog.coding.for.Logic.synthesis.(2003)	17.471	19.03.2011	04.08.2012	504	34,7
Wiley - Convergence Technologies for 3G Networks - IP, UMTS, EGPRS and	13.403	14.06.2011	04.08.2012	417	32,1
WileyCIA P1 All Q&A	12.752	24.09.2010	04.08.2012	680	18,8
wiley.wireless.lan.radios.dec.2007	10.468	08.02.2011	04.08.2012	543	19,3
Wiley CIA 2006 v1 Page 1 of 262 Question: V2C3-0001 An internal audit t	9.837	08.06.2011	04.08.2012	423	23,3
4-Accrual Accounting Concepts	8.924	26.07.2012	04.08.2012	9	991,6
Wiley Simulation the Practice of Model Development and Use	8.553	04.01.2012	04.08.2012	213	40,2
Wiley Simulation the Practice of Model Development and Use	8.516	09.05.2011	04.08.2012	453	18,8
Wiley - Statistical Digital Signal Processing and Modeling	8.479	27.11.2011	04.08.2012	251	33,8
Bannerjee - Redox Biochemistry (Wiley, 2007)	8.281	13.06.2012	04.08.2012	52	159,3
Unified Theory of Concrete Structures - 0470688742 - Wiley	8.271	23.10.2010	04.08.2012	651	12,7
Wiley HTML5, Your Visual Blueprint for Designing Rich Web Pages and Ap	7.953	19.08.2010	04.08.2012	716	11,1
Wiley IP Multicast With Applications to IPTV and Mobile	7.545	08.01.2010	04.08.2012	939	8,0
Wiley IP Multicast With Applications to IPTV and Mobile	7.523	25.04.2011	04.08.2012	467	16,1
John.wiley.and.Sons.wcdmA for UMTS HSPA Evolution and LTE	7.005	12.05.2012	04.08.2012	84	83,4
Ebooksclub.org Handbook of Monte Carlo Methods Wiley Series in Probab	6.731	26.10.2011	04.08.2012	283	23,8
Wiley - Information Technology Project Management	6.533	05.07.2012	04.08.2012	30	217,8
Wiley Finance,. Practical Risk Management.[2003.ISBN0470849673]	6.506	15.12.2011	04.08.2012	233	27,9

The table contains 400 books. Using the search engine of Scribd we could identify more than 1000 titles (as far as the output of the search engine is limited to 500 results we could not identify all possible title incl. copies, so that the number 1.000 is assumed to be a lower limit). For the calculation of the lost sales, we are using the following formula:

$$\text{Lost revenues} = \text{number of downloads} * \text{price of the book} * \text{replacement rate}$$

The only unknown parameter is the replacement rate that defines how many books are downloaded to replace one actual sale. The music industry assumes a replacement rate of 30%. That seems high in our opinion. We assume in the calculation a replacement rate of 1%. That means on 100 downloads one sale is prevented – a relatively tiny number, smaller than the usual rate of losses by shop lifting! If you assume a smaller replacement rate it wouldn't make sense to talk about a piracy problem at all.

Based on these input number one will get for this example the following results

Total Downloads	538.533
Replacement rate (assumed, definitely higher)	1%
Average price	\$50
Losses (of 400 books)	\$269.267
Loss per Title	\$1.346
Titels at scribd (estimated, minimum visible)	1.000
Total losses (estimation, 1000 books available)	\$1.346.333

This means, that the 1000 ebooks we have assumed as the lower limit of the available books of Wiley (at scribd.com only) have generated revenue losses of about 1.4 million USD in the time those books have been online. This number defines the order of magnitude of the piracy problem. Besides of 1.000 books of Wiley in this example, there are at least 20.000 illegal ebook copies available on different platforms and communities with an unknown number of copies per single book. The replacement rate of 1% is also a *very* conservative assumption – next to nothing! Even that small number generates revenue losses larger than a million. The actual revenue losses can be assumed to reach multiple millions of Dollars per year and per (STM) publisher.

Removing the illegal uploads from Scribd would stop those losses, at least those generated via scribd.com. Wiley is one large STM publisher. As one can see on the table of the Library Genesis given above there are others, facing the same or a similar problem. All together one can assume that the revenue losses of all publishers caused by illegal copies exceed the billion Dollars per year.

The numbers we have found are in agreement with those of [10] although we have used a different methodology. The statement of [10] that there are (4/2011) about 1 million links seems very conservative. As one can see above, one single site offers more than 1.2 million books. There are more sites than that one! We assume the number of links leading to illegal book copies is 1-2 orders of magnitude larger – and so is the ebook piracy problem!

At present it seems, that publishers ignore the problems of piracy. It was in our intention to demonstrate in a clear way the scales of the revenue losses. We have chosen one STM publisher (Wiley) only as one example of a calculation of revenue losses. The problem does concern nearly all of those publishers.

Tempus fugit! Each day without measures against the illegal distribution of ebooks will cost the publishers (and so the authors, book shop owners etc.) more money. It was our attempt to show, how much it is and how this number can be estimated.

References

- [1] Maria-Leena Räihälä, morgenvogel.net
- [2] http://abuse-search.com/Gutenberg3.0-Ebook_Piraterie_in_Deutschland.pdf
- [3] <http://abuse-search.com/Gutenberg-31.pdf>
- [4] <http://abuse-search.com/Piracy-for-STM-publishers.pdf>
- [5] <http://www.alex.com>
- [6] <http://www.zeit.de/kultur/literatur/2012-06/buchmarkt-ebook-verkaufszahlen>
- [7] <http://gutenberg.spiegel.de/>
- [8] <http://www.bgr.com/2012/07/12/ipad-sales-estimates-q2-2012/>
- [9] http://www.buchreport.de/nachrichten/handel/handel_nachricht/datum/2011/07/25/schwachbruestige-buecher.htm
- [10] <http://attributor.com/data/pdf/infographic05-2011.pdf>