

### 入门学习

---

1. 信息检索导论学习笔记 ( 1-10 ) [[http://www.cnblogs.com/god\\_bless\\_you/archive/2012/08/10/2631975.html](http://www.cnblogs.com/god_bless_you/archive/2012/08/10/2631975.html)]
2. TF-IDF与余弦相似性的应用 ( 一 ) : 自动提取关键词 [<http://www.ruanyifeng.com/blog/2013/03/tf-idf.html>]
3. 信息检索——bm25算法原理及实现 [<http://www.jianshu.com/p/1e498888f505>]
4. 数学之美 吴军 [<https://book.douban.com/subject/10750155/>]
5. Introduction to Information Retrieval. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. [<https://nlp.stanford.edu/IR-book/>]

### 进阶文章

---

#### 1960

---

1. [Maron&Kuhn 1960] On relevance, probabilistic indexing and information retrieval. M. E. Maron and J. L. Kuhns. Journal of the ACM, 1960.  
[<https://dl.acm.org/citation.cfm?id=321035>]

#### 1975

---

1. [Salton&Yang&Yu 1975] A theory of term importance in automatic text analysis. G. Salton, C.S. Yang and C. T. Yu. Journal of the American Society for Information Science, 1975. References  
[<https://eric.ed.gov/?id=ED096987>]
2. [Harter 1975] A probabilistic approach to automatic keyword indexing. S. P. Harter. Journal of the American Society for Information Science, 1975.  
[[https://www.researchgate.net/publication/240325312\\_A\\_probabilistic\\_approach\\_to\\_automatic\\_keyword\\_indexing\\_Part\\_I\\_On\\_the\\_Distribution\\_of\\_Terms\\_in\\_Indexing](https://www.researchgate.net/publication/240325312_A_probabilistic_approach_to_automatic_keyword_indexing_Part_I_On_the_Distribution_of_Terms_in_Indexing)]

#### 1976

---

1. [Robertson&Sparck Jones 1976] Relevance weighting of search terms. S. Robertson and K. Sparck Jones. Journal of the American Society for Information Science, 1976.  
[<http://www.staff.city.ac.uk/~sb317/papers/RSJ76.pdf>]

#### 1977

---

1. [van Rijsbergen 1977] A theoretical basis for the use of cooccurrence data in information retrieval. C. J. van Rijsbergen. Journal of Documentaion, 1977.  
[<http://www.emeraldinsight.com/doi/pdfplus/10.1108/eb026637>]

#### 1994

---

1. [Robertson&Walker 1994] Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. S. E. Robertson and S. Walker. SIGIR 1994.  
[<https://dl.acm.org/citation.cfm?id=188561>]

#### 1996

---

1. [Singhal et al. 1996] Pivoted document length normalization. A. Singhal, C. Buckley and M. Mitra. SIGIR 1996.  
[<https://dl.acm.org/citation.cfm?doid=243199.243206>]
2. [Singhal et al. 1996] Pivoted document length normalization. A. Singhal, C. Buckley and M. Mitra. SIGIR 1996.  
[<https://dl.acm.org/citation.cfm?doid=243199.243206>]

#### 1998

---

1. [Ponte&Croft 1998] A language modeling approach to information retrieval. J. Ponte and W. B. Croft. SIGIR 1998.  
[<https://dl.acm.org/citation.cfm?doid=290941.291008>]

#### 2001

---

1. [Zhai&Lafferty 2001] A study of smoothing methods for language models applied to ad hoc information retrieval. C. Zhai and J. Lafferty. SIGIR 2001.

[<https://dl.acm.org/citation.cfm?id=384019>]

2. [Lavrenko&Crof 2001] Relevance-based language models. V. Lavrenko and B. Crof. SIGIR 2001.

[<https://dl.acm.org/citation.cfm?id=383972>]

---

## 2004

1. [Kurland&Lee 2004] Corpus structure, language models, and ad hoc information retrieval. O. Kurland and L. Lee. SIGIR 2004.

[<https://dl.acm.org/citation.cfm?id=1009027>]

---

## 2013

1. [Huang et al. 2013] Huang et al. Learning deep structured semantic models for web search using clickthrough data, 2013 CIKM.

[[https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/cikm2013\\_DSSM\\_fullversion.pdf](https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/cikm2013_DSSM_fullversion.pdf)]

2. [Lu et al. 2013] Lu Z, Li H. A deep architecture for matching short texts //NIPS 2013: 1367-1375

[[https://cis.temple.edu/~yuhong/teach/2014\\_spring/papers/NIPS2013\\_DeepText.pdf](https://cis.temple.edu/~yuhong/teach/2014_spring/papers/NIPS2013_DeepText.pdf)]

---

## 2014

1. [Montúfar et al. 2014] Montúfar, Pascanu, Cho and Bengio. On the number of linear regions of deep neural networks NIPS 2014

[[http://www.mis.mpg.de/preprints/2014/preprint2014\\_73.pdf](http://www.mis.mpg.de/preprints/2014/preprint2014_73.pdf)]

2. [Shen et al. 2014] Shen et al. A latent semantic model with convolutional-pooling structure for information retrieval. 2014 CIKM

[<https://dl.acm.org/citation.cfm?id=2661935>]

3. [Denil et al. 2014] Denil M, Demiraj A, Kalchbrenner N, et al. Modeling, visualising and summarising documents with a single convolutional neural network[J]. arXiv preprint arXiv: 1406.3830, 2014.

[<https://arxiv.org/abs/1406.3830>]

4. [Kalchbrenner et al. 2014] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modeling sentences[J]. arXiv preprint: 1404.2188, 2014.

[<https://wenku.baidu.com/view/6016f5a55fbfc77da369b15f.html>]

5. [Kim et al. 2014] Kim Y. Convolutional neural networks for sentence classification [J]. arXiv preprint: 1408.5882, 2014.

[<https://arxiv.org/abs/1408.5882>]

6. [Hu et al. 2014] Hu B, Lu Z, Li H, et al. Convolutional neural network architectures for matching natural language sentences//NIPS 2014: 2042-2050

[<https://arxiv.org/abs/1503.03244>]

---

## 2015

1. [He et al. 2015] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] //CVPR 2016: 770- 778.

[<https://arxiv.org/abs/1512.03385>]

2. [Lin et al. 2015] Lin M, Zhengdong L, Lifeng S et al. Multimodal Convolutional Neural Network for matching Image and sentence//CVPR

[<https://arxiv.org/abs/1504.06063>]

3. [Alessandro et al. 2015] Alessandro S, Yoshua B, Hossein V. et al. A hierarchical recurrent encoder-decoder for generative context-aware query suggestion[c]//CIKM 2015: 553-562

[<https://dl.acm.org/citation.cfm?doid=2806416.2806493>]

4. [Mitra et al. 2015] Mitra B, Craswell N. Query auto-completion for rare prefixes[c]//CIKM 2015: 1755-1758

[<https://dl.acm.org/citation.cfm?id=2806599>]

5. [Severyn and Moschitti 2015] Severyn A, Moschitti A. Learning to rank short text pairs with convolutional deep neural networks [C]// SIGIR 2015: 373-382.

[[http://casa.disi.unitn.it/~moschitti/since2013/2015\\_SIGIR\\_Severyn\\_LearningRankShort.pdf](http://casa.disi.unitn.it/~moschitti/since2013/2015_SIGIR_Severyn_LearningRankShort.pdf)]

6. [Tai et al. 2015] Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks[J]. arXiv preprint arXiv: 1503.00075, 2015.

[<https://arxiv.org/abs/1503.00075>]

7. [Zhou et al. 2015] Zhao H, Lu Z. Poupart P. Self-adaptive Hierarchical Sentence Model[C] ///IJCAI. 2015: 4069- 4076

[<https://arxiv.org/abs/1504.05070>]

---

## 2016

1. [Larsson et al. 2016] Larsson G, Maire M, Shakhnarovich G. Fractalnet: Ultra-deep neural networks without residuals [J]// axXiv preprint: 1605.07648, 2016.

[<https://arxiv.org/abs/1605.07648>]

2. [Szegedy et al. 2016 ] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C] //CVPR 2015: 1-9.

[<https://arxiv.org/abs/1409.4842>]

3. [Palangi et al. 2016] Palangi, et al. Deep sentence embedding using long short-term memory networks TASLP 2016  
[<https://arxiv.org/abs/1502.06922>]
4. [Wan et al. 2016] Wan S, Lan Y, Guo J, et al. A deep architecture for semantic matching with multiple positional sentence representations//Proceedings of the 30th AAAI Conference on Artificial Intelligence . Phoenix, USA, 2016: 2835-2841.  
[<https://arxiv.org/abs/1511.08277>]
5. [Guo et al. 2016] Jiafeng G, Yixing F, Qingyao A et al. A Deep Relevance Matching Model for ad-hoc retrieval//CIKM 2016: 55-64  
[[https://www.researchgate.net/publication/310823413\\_A\\_Deep\\_Relevance\\_Matching\\_Model\\_for\\_Ad-hoc\\_Retrieval](https://www.researchgate.net/publication/310823413_A_Deep_Relevance_Matching_Model_for_Ad-hoc_Retrieval)]
6. [Yan et al. 2016] Yan R, Song Y, Wu Learning to respond with deep neural networks for retrieval-based humancomputer conversation system//SIGIR 2016:55-64  
[<https://dl.acm.org/citation.cfm?id=2911542>]
7. [Zhou et al. 2016] Zhou X, Dong D, Wu H, et al. Multi-view Reponse Selection for Human-Computer Conversation [C] //EMNLP 2016: 372-381.  
[<http://wing.comp.nus.edu.sg/~antho/D/D16/D16-1036.pdf>]
8. [Wan et al. 2016] Wan S, Lan Y, Xu J, et al. Match-SRNN: Modeling the recursive matching structure with spatial rnn[J]. IJCAI 2016  
[<https://arxiv.org/abs/1604.04378>]
9. [Liang et al. 2016] Liang P, Yanyan L, Jiafeng G et al. Text Matching as Image Recognition//AAAI 2016: 2793-2799  
[<http://arxiv.org/abs/1602.06359>]
10. [Long et al. 2016] Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng, Modeling Document Novelty with Neural Tensor Network for Search Result Diversification, the 39th Annual ACM SIGIR Conference, Pisa, Italy (SIGIR 2016)  
[<https://dl.acm.org/citation.cfm?id=2911498>]

2017

1. [Mitra et al. 2017] Mitra B, Diaz F, Craswell N. Learning to Match using Local and Distributed Representations of Text for Web Search//WWW 2017: 1291-1299  
[<https://dl.acm.org/citation.cfm?doid=3038912.3052579>]

## Tutorial

1. Neural Networks for Information Retrieval by Tom Kenter<sup>1</sup> Alexey Borisov<sup>1,2</sup> Christophe Van Gysel<sup>1</sup> Mostafa Dehghani<sup>1</sup> Maarten de Rijke<sup>1</sup> Bhaskar Mitra Full-day tutorial at SIGIR 2017 on August 7, 2017 [<http://nn4ir.com/slides/NN4IR.pdf>]
2. Machine Learning for Information Retrieval 李航 华为诺亚实验室 [[http://www.hangli-hl.com/uploads/3/4/4/6/34465961/machine\\_learning\\_for\\_information\\_retrieval\\_kyoto.pdf](http://www.hangli-hl.com/uploads/3/4/4/6/34465961/machine_learning_for_information_retrieval_kyoto.pdf)]
3. 深度学习与信息检索 第一节 基于分布式表示的信息检索 第二节 基于深层匹配的信息检索 [<http://cips-upload.bj.bcebos.com/2017/ssatt2017/ATT2017-IRI.pdf>] [<http://cips-upload.bj.bcebos.com/2017/ssatt2017/ATT2017-IRII.pdf>]

## 综述

1. Information Retrieval Surveys [<http://www.umi.acs.umd.edu/~oard/teaching/734/fall14/irsurveys.html>]
2. Neural Networks for Information Retrieval 微软 Bhaskar Mitra Nick Craswell [<https://arxiv.org/pdf/1705.01509.pdf>]
3. 深度文本匹配综述 庞亮, 兰艳艳 徐君, 郭嘉丰, 万圣贤, 程学旗 中科院计算所 [<http://cjc.ict.ac.cn/online/onlinepaper/pl-201745181647.pdf>]

## 视频教程

1. 信息检索 武汉大学课程 [<http://www.icourse163.org/course/WHU-29001#/info>]

## 代码

1. Apache Lucene Lucene是apache软件基金会4 jakarta项目组的一个子项目, 是一个开放源代码的全文检索引擎工具包, 但它不是一个完整的全文检索引擎, 而是一个全文检索引擎的架构, 提供了完整的查询引擎和索引引擎, 部分文本分析引擎。Lucene的目的是为软件开发人员提供一个简单易用的工具包, 以方便的在目标系统中实现全文检索的功能, 或者是以此为基础建立起完整的全文检索引擎。 [<https://lucene.apache.org/core/>]
2. The Lemur Project Lemur ( 狐猴 ) 系统是CMU和UMass联合推出的一个用于自然语言模型和信息检索研究的系统。在这个系统上可以实现基于自然语言模型和传统的向量空间模型以及Okapi的ad hoc或者分布式检索, 可以使用结构化查询、跨语言检索、过滤、聚类。

## 领域专家

1. Christopher D .Manning 斯坦福大学计算机与语言学教授, 他于1995年在斯坦福大学获得博士学位。在执教斯坦福大学之前他先后在卡耐基梅隆大学 ( CMU ) 和悉尼大学任教。Manning的研究理想是使计算机能够智能地处理、理解并生成人类语言。他的研究专注于利用机器学习方法处理计算语言学问题, 其中包括句法分析、计算语义和语用学、文本推理、机器翻译及面向自然语言处理的递归深度学习等, 特别是其引领了深度学习在自然语言处理领域的研究。Manning教授是ACM Fellow, AAAI Fellow 和ACLFellow, 目前任ACL主席。Manning教授所著的多部书籍都成为了权威经典教材, 例如《统计自然语言处理基础》、《信息检索导论》等。 <http://nlp.stanford.edu/~manning/>

2. Susan Dumais 微软雷德蒙研究院首席研究员，人机交互和信息检索方面的国际知名专家。美国计算机学会信息检索专委会前主席，国际信息检索大会（SIGIR）2006的程序委员会主席，美国计算机学会院士，美国国家工程院院士，格拉德•撒尔顿奖（Gerard Salton Award，信息检索领域最高奖项）得主。[<http://susandumais.com/>]
3. Ricardo Baeza-Yates 雅虎欧洲、中东、拉美研究院副总裁，信息检索和数据挖掘方面的国际知名专家。国际信息检索大会2011年的程序委员会主席，美国计算机学会院士，国际电子电气工程师学会院士，信息检索领域经典著作《现代信息检索》(Modern Information Retrieval)的作者。[<http://www.baeza.cl/>]
4. Stephen Robertson 微软剑桥研究院高级研究员，信息检索模型方面国际公认的先驱，著名的信息检索模型Okapi BM25的发明人。英国剑桥哥顿学院院士，格拉德•撒尔顿奖得主。[<http://www.nicholsons.com.au/team/stephen-robertson>]
5. 李航 前华为技术有限公司诺亚方舟实验室主任，现已加入今日头条[<http://www.hangli-hl.com/index.html>]
6. 刘铁岩 博士 微软亚洲研究院研究主管。信息检索和排序学习领域的国际知名专家，曾获得国际信息检索大会最佳论文奖和国际图像通信和视觉表达期刊的最高引用论文奖。多次担任国际信息检索大会(SIGIR)和国际互联网大会(WWW)的领域主席，现任国际信息检索期刊(IR Journal)编委，美国计算机学会信息系统期刊(TOIS)副主编，国际电子电气工程师学会高级会员。[<https://www.microsoft.com/en-us/research/people/tyliu/>]
7. 刘挺 哈尔滨工业大学教授，社会计算与信息检索研究中心主任。多次担任国家863重点项目总体组专家、973项目专家组成员、基金委评审专家，入选科技部中青年科技创新领军人才。中国计算机学会（CCF）理事、CCF哈尔滨主席；中国中文信息学会常务理事、社交媒体处理专委会主任；曾任国际顶级会议ACL、EMNLP领域主席，多次担任全国学术会议（SMP、CCIR）的大会主席、程序委员会主席。主要研究方向为自然语言处理和社会计算，是国家973课题、国家自然科学基金重点项目的负责人。曾获国家科技进步二等奖、省科技进步一等奖、钱伟长中文信息处理科学技术一等奖。[<http://ir.hit.edu.cn/>]
8. 翟成祥 教授（Prof. Chengxiang Zhai）美国伊利诺伊大学香槟分校副教授，语言模型方面的国际知名专家。国际信息检索大会2009年的程序委员会主席，美国计算机学会杰出科学家。
9. 王斌 中科院信工所 研究员，博导，中国科学院信息工程研究所第二研究室信息检索课题组组长，研究方向为信息检索与自然语言处理。[<http://peopleucas.edu.cn/~wangbin/>]
10. Jiafeng Guo（郭嘉丰）郭嘉丰，博士，现任中科院计算技术研究所研究员，中国科学院大学岗位教授，长期从事信息检索与数据挖掘方向的基础研究，研究方向包括用户查询理解、排序学习、文本建模、神经网络搜索等。在本领域国际重要学术会议与期刊上发表学术论文80余篇（如TKDE、SIGIR、AAAI、CIKM等），Google Scholar累计引用1800余次，获得ACM CIKM 2011 最佳论文奖，ACM SIGIR 2012最佳学生论文奖，2012年国家科技进步二等奖、2012年中国中文信息学会“钱伟长中文信息处理科学技术奖——汉王青年创新奖”一等奖，2016年中科院青年促进会优秀会员等[<http://www.bigdatalab.ac.cn/~gjf/>]

初步版本，水平有限，有错误或者不完善的地方，欢迎大家提建议和补充，会一直保持更新，敬请关注<http://www.zhuanzhi.ai> 和关注**专知**公众号，获取第一手AI相关知识

