分布式智能博客的研究

## 基本概念的概述

博客：英文名为[Blogger](https://baike.baidu.com/item/Blogger)，为Web Log的混成词。它的正式名称为网络日记；又[音译](https://baike.baidu.com/item/%E9%9F%B3%E8%AF%91" \t "_blank)为[部落格](https://baike.baidu.com/item/%E9%83%A8%E8%90%BD%E6%A0%BC)或部落阁等，是使用特定的软件，在网络上出版、发表和张贴个人文章的人，或者是一种通常由个人管理、不定期张贴新的文章的[网站](https://baike.baidu.com/item/%E7%BD%91%E7%AB%99/155722)。

分布式【15】：分布式系统（distributed system）是建立在网络之上的[软件](https://baike.baidu.com/item/%E8%BD%AF%E4%BB%B6" \t "_blank)系统。正是因为[软件](https://baike.baidu.com/item/%E8%BD%AF%E4%BB%B6/12053)的特性，所以分布式系统具有高度的[内聚性](https://baike.baidu.com/item/%E5%86%85%E8%81%9A%E6%80%A7/4973441)和透明性。因此，网络和分布式系统之间的区别更多的在于高层[软件](https://baike.baidu.com/item/%E8%BD%AF%E4%BB%B6/12053)（特别是[操作系统](https://baike.baidu.com/item/%E6%93%8D%E4%BD%9C%E7%B3%BB%E7%BB%9F/192" \t "_blank)），而不是硬件。

## 博客系统的意义概述

国内外都有大量的博客系统，比如国内的博客园、CSDN等等，国外的WordPress、**Movable Type等等。这些博客系统的出现以及流行，都有其存在的意义。博客有其独特的文化和社会意义。进10年来，博**客已在整个世界范围尤在中国取得长足和飞速发展。 博客改变了传统的书写方式,给出版和传媒带来新的观念和模式,也带来冲击,开创了知识和思想自由传播的新时代。 正如方兴东的《博客宣言》指出的:“作为信息时代的麦哲伦,博客的出现使互联网世界第一次有了知识积累和文化指向。 使人类由粗放的数字化生存,过渡为个人的精确的目录式生存。 如同麦哲伦的航海日志一样,博客将工作、生活和学习融为一体,通过博客日志,将日常的思想精华及时记录并发布,萃取并联接全球最有价值、最相关、最有意思的信息与资源。使更多的知识工作者能够零距离、零壁垒地汲取这些最鲜活的思想。 对于中国这样一个知识希缺的国度,博客的繁荣,具有更特别的意义。 博客在中国不仅是航海者,还是盗火者。在汉语世界中的自娱自乐,最终将会使中国走向狭隘,而对新知的探索之旅越加充满荆棘。我们希望,博客文化能引领中国向知识社会转型,博客关怀能开启一个负责的时代。”的确,博客这种新的知识和思想运作模式对中国社会的影响是深刻而深远的,它深入到了个体心理、精神等层面,从各个方面对个体和社会施予“纹”与“化”之功。下文透过几件比较有影响的大事,探讨博客对社会的影响,揭示博客对中国社会的积极影响【6】。另外博客为青少年提供了一个展示真实自我、发挥个人特性的渠道，同时,也在某种程度上建立了一种新的文化符号。但是也有不少研究者认为，过多地沉溺于网络和博客，会减少与家人和朋友共度的时光,导致心理健康程度下降.不可否认,由于网络的匿名性等特点，很难保证网络平台不被一些心怀不轨的人所利用。因此，博客要真正承担起健康积极的文化影响职责，还有赖于健康有序的网络文化,甚至媒介文化的培育，而在这点上，任重而道远【7】。

## 国内外对博客系统的研究现状

由于博客系统的有益性，在国内外都对其进行了大量的研究，力求寻到博客系统的进步和发展。如研究博客留存率的Imrul Kayes【1】等人（他们询问是什么因素促使博客作者继续通过贡献内容(例如，帖子、评论)来参与社区。从一个流行的社区博客平台“Blogster”抓取了一个博客作者简介样本(贡献了91%的帖子)。结果显示，男性和年长(资深)博主在社区中面临较少的限制和更多的机会，比其他人更容易被保留。其他博主通过隐性的(阅读文章)和显性的(写评论)互动对这些留存的博主给予高度关注。他们还发现，如果一个博主的朋友也有更高的用户留存率，那么这个博主的朋友留存率就会更高，而且强大的社会联系可以减少两个博主朋友之间的用户留存不平衡。然而，他们发现一个博主的网络时代(例如，她加入多久)对她的留存率没有影响），还有他们的另外一篇研究【2】（社区博客是一种发表日常日志、表达观点或想法以及分享知识的媒体。博客对市场营销、塑造公众舆论、从草根的角度向世界传达重大事件都有很大的影响。然而，在线博客的流动率非常高，大多数最初加入并开始为社区做出贡献的人，都无法长期贡献。他们想知道是什么因素促使博主继续通过贡献内容(例如，帖子、评论)来参与社区。他们从一个流行的社区博客平台“Blogster”上抓取了一个博客个人资料样本。这些博主在社区中贡献了91%的帖子。他们推导了一组与博客用户留存率相关的可靠变量，并从这些变量中建立了预测模型。他们的研究结果显示，男性和年长的博主在社区中面临较少的限制和更多的机会，比其他人更容易被保留。其他博主通过隐性(阅读帖子)和显性(写评论)的互动对这些留存的博主给予高度关注。他们还发现，如果一个博主的朋友也有更高的用户留存率，那么她的朋友留存率就会更高，而强大的社会联系可以减少两个博主朋友之间的用户留存不平衡。然而，他们发现一个博主的网络时代(例如，她加入多久)对她的留存率没有影响。他们的工作对博客作者的社会行为文献有理论意义，对潜在的社区博客平台开发者有实际意义）；研究博客本身搜索技术的João Simões等人【3】（他们介绍了目前存在的博客搜索模型和博客搜索引擎，并提出了一个搜索博客的模型。这种模式使用了几种关键字搜索技术，也基于几个api，这些api可以从博客平台上获取最新的帖子，以及RSS订阅，在博客内容更新的战略性网站上仍然使用网络爬虫。该模型的目的是创建一个能够捕获大量博客并存储其内容的解决方案）；如研究博客数据的Adewale  Obadimu等人【4】（社交媒体已经发展成为表达观点、分享信息和塑造话语的场所。个人利用社会媒体作为平台来动员、协调和开展网络运动，从提高对疾病或失调的认识，到威胁民主原则和制度的越轨行为。博客圈持续崛起，为内容框架提供了有效的媒介。在字数上没有限制的情况下，许多人使用博客来设置叙述，然后使用其他的社交媒体渠道，如Twitter和Facebook来引导他们的受众访问他们的博客。与其他社交媒体渠道相比，博客内容没有结构化，很难收集。博客监测和分析对于社会学家、政治学家、传播学研究者、新闻工作者和信息科学家检查事件是非常有用的。为此，他们推出了博客追踪工具，旨在探索博客圈并获得对各种事件的见解。博客追踪者可以帮助识别领先的信息参与者、有影响力的博客作者、流行的和新兴的趋势，评估语气、情绪和观点，提取实体，并分析他们的网络）；研究博客中的信息共享对大学生博客使用模式及性别的影响的Ji-Hyun Kim【14】等人（本研究以25项陈述来衡量学生使用博客的情况、性别差异、态度及对使用博客的满意度。这25个陈述分为四个领域:信息共享、社会沟通、管理和负面影响。研究结果表明，学生主要通过博客来分享和传播信息。与同学联系和提供个人生活的细节对学生来说也很重要）；还有国内的张宝军等人研究的结合云存储的博客系统【5】(为解决新一代博客系统海量信息的存储问题,结合云存储技术,提出了一种新的博客系统架构Blog Cloud。该架构以分布式存储技术为核心,避免了集中式存储的性能瓶颈问题,可扩展性高;采用半分布式P2P网络拓扑结构,能快速定位网络中的存储资源;只将稳定节点作为存储节点,避免了不稳定节点带来的网络波动问题;遵循就近存储原则,同时在客户端缓存文件,减少了网络传输;允许用户自定义文件分块的大小,对大的文件可分块并行传输,提高了文件传输的速度,对小的文件则不用分块,节省了文件分块、合并的开销;具备数据冗余备份功能,在网络中多个存储节点保留文件副本,并实行异地备份,增强了数据的安全性和可靠性。在虚拟机上对Blog Cloud和ZSWIN博客系统进行比较测试,结果显示:Blog Cloud的吞吐量明显高于ZSWIN;将不稳定节点作为存储节点会降低Blog Cloud的性能;在存储节点和索引节点减少的情况下Blog Cloud仍然能够稳定运行,可靠性较高。结果表明,Blog Cloud架构能够满足新一代博客系统的存储要求)。

## 博客系统的主要实现步骤

以现有的项目开发步骤为例，【8】

1. 可行性研究

可行性研究（Feasibility Study），是指在调查的基础上，通过市场分析、[技术分析](https://baike.baidu.com/item/%E6%8A%80%E6%9C%AF%E5%88%86%E6%9E%90" \t "_blank)、财务分析和[国民经济](https://baike.baidu.com/item/%E5%9B%BD%E6%B0%91%E7%BB%8F%E6%B5%8E)分析，对各种投资项目的技术可行性与经济合理性进行的综合评价。可行性研究的基本任务，是对新建或[改建项目](https://baike.baidu.com/item/%E6%94%B9%E5%BB%BA%E9%A1%B9%E7%9B%AE)的主要问题，从技术[经济](https://baike.baidu.com/item/%E7%BB%8F%E6%B5%8E)角度进行全面的分析研究，并对其投产后的经济效果进行预测，在既定的范围内进行方案论证的选择，以便最合理地利用资源，达到预定的社会效益和[经济效益](https://baike.baidu.com/item/%E7%BB%8F%E6%B5%8E%E6%95%88%E7%9B%8A)

1. 需求分析【13】

需求分析也称为软件需求分析、系统需求分析或需求分析工程等，是开发人员经过深入细致的调研和分析，准确理解用户和项目的功能、性能、可靠性等具体要求，将用户非形式的需求表述转化为完整的需求定义，从而确定系统必须做什么的过程。需求分析，捕获人们想要什么或需要什么，而不是确定他们喜欢什么或不喜欢什么，这是许多传统工作的重点。

1. 总体设计

总体设计的基本目的是回答“概括地说，系统应该如何实现”。这个阶段的工作划分一个大的功能模块，不需要非常细致，以及找寻各个功能模块之间的关系。

1. 详细设计

详细设计，是[软件工程](https://baike.baidu.com/item/%E8%BD%AF%E4%BB%B6%E5%B7%A5%E7%A8%8B" \t "_blank)中软件开发的一个步骤，就是对[概要设计](https://baike.baidu.com/item/%E6%A6%82%E8%A6%81%E8%AE%BE%E8%AE%A1/9827718)的一个细化，就是详细设计每个模块实现算法，所需的局部结构。在详细设计阶段，主要是通过需求分析的结果，设计出满足用户需求的软件系统[产品](https://baike.baidu.com/item/%E4%BA%A7%E5%93%81/105875)。

1. 编码

编写代码，开发。

1. 测试【9】

包括单元测试，集成测试，系统测试等，实现对开发出来的系统进行全方面的测试，最后可以正常移交给用户。

1. 维护【12】

在软件项目的开发中扮演着关键的角色，对开发的系统进行持续性的维护升级。

## 博客系统功能要解决的关键问题

1. 博客文章的内容分析【10】

Blog作为继E-mail、BBS和ICQ之后出现的一种新型网络交流工具,逐渐在传媒组织、经济组织、事业组织、政府组织的现代化运作过程中得到广泛应用.但是由于其内容的不确定性，导致博客可能出现一些不良的言论影响，故需要对整个博客内容进行分析、限制和审核。

1. 博客推荐【11】

Tumblr等热门微博网站作为一个内容分享平台吸引了数亿用户，用户可以以帖子的形式创建丰富的内容，并与关注他们的其他用户分享。由于在这些服务上创建的帖子数量庞大，一个重要的任务是为用户提供高质量的博客推荐。除了以追随者图为主要数据源的传统推荐系统设置外，还可以利用用户和博客的附加侧信息，如用户活动(如like和reblog)和丰富的内容(如文本和图像)，以提高推荐性能。在本文中，我们提出了一种新的增强归纳矩阵补全方法(BIMC)用于博客推荐。BIMC是一种可加性的低秩用户博客偏好模型，由两部分组成;一个组件捕获跟随关系的低级结构，另一个使用侧信息捕获潜在结构。我们的模型公式结合了最近提出的归纳矩阵补全(IMC)模型(用于侧信息)和标准矩阵补全(MC)模型(用于低秩结构)的能力。此外，我们利用最近开发的深度学习技术来获取包含在BIMC中的文本和图像的丰富语义特征表示。在Tumblr上的大规模真实数据集上的实验证明了该方法的有效性

## 总结

本文总结了博客系统的意义以及实现分布式智能博客系统主要的步骤，分析了当前系统在国内外的研究现状。由于博客的有益性使得其得了到广泛的应用。目前我的工作就是开发出这样一个博客系统，来满足用户的某些需求，以及不断模拟数据，进行博客推荐的验证和优化。

## 参考文献

[1] Did You Blog Yesterday? Retention in Community Blogs Imrul Kayes, Xiang Zuo, Da Wang ,Jacob Chakareski 2014

[3] WISE Blogs:A Special Blog Search Engine João Simões, Frederico Azeiteiro,Jorge Bernardino 2015

[2] To Blog or Not to Blog: Characterizing and Predicting Retention in Community Blogs Imrul Kayes, Xiang Zuo, Da Wang ,Jacob Chakareski 2014

[4] Blog data analytics using blogtrackers Adewale Obadimu, Muhammad Nihal Hussain, Nitin Agarwal 2019

[5] 结合云存储的新一代博客系统架构 张宝军，潘瑞芳 2015

[6] 博客的文化与社会意义 马云驰 2010

[7] 青少年博客的社会文化意义研究 孙信茹，杨星星 2010

[8] 软件工程导论 （第六版） 张海藩，牟永敏 2013

[9] 软件测试技术 曲朝阳，刘志颖，杨杰明，刘迪 2015

[10] 组织系统博客文化的技术和内容分析 武彬，黄杰 2007

[11] Tumblr Blog Recommendation with Boosted Inductive Matrix Completion Donghyuk Shin，Suleyman Cetintas，Suleyman Cetintas，Suleyman Cetintas 2015

[12] Build system maintenance Shane McIntosh 2011

[13] Textual demand analysis: detection of users' wants and needs from opinions Hiroshi Kanayama，Tetsuya Nasukawa 2008

[14] Information sharing in blog: exploring college students' blog usage pattern and gender difference Ji-Hyun Kim，Hwan Jin Choi 2010

[15] George Coulouris Distributed SystemsConcepts and Design Fith Edition 2011