第一图书网, tushu007.com

<<预测性文本挖掘基础>>

图书基本信息

书名：<<预测性文本挖掘基础>>

13位ISBN编号：9787560544274

10位ISBN编号：7560544274

出版时间：2012-10

出版时间：西安交通大学出版社

作者：[美] 绍洛姆•韦斯(Sholom M.Weiss) [澳]尼亭•因杜尔亚（Nitin Indurkhya）[美]张潼(Tong zhang),赵仲孟 侯， 迪 译

页数：234

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：http://www.tushu007.com

第一图书网, tushu007.com

<<预测性文本挖掘基础>>

内容概要

本书是文本挖掘这个迅速发展领域的入门性教材和指南。
作为入门型读物，论述由浅入深、理论结合实践、语言风趣、样例详实，详细地分析和总结了该领域的研究现状及未来发展。
深入地讨论了文档分类、信息检索、聚类与组织文档、信息提取、基于Web的数据源、预测与评价等方面提出的问题。
作为互联网和云计算蓬勃发展时代，为关注海量非结构化信息处理的学习者了解文本挖掘研究前沿打开了一扇窗口。

特别邀请知名的台湾辅仁大学统计资讯系谢邦昌教授倾情作序，评价“这本书为文本挖掘入门最佳的书”。

本书原版是计算机科学方向的一本国外教材，译本适用于计算机专业本科生和研究生作为关键教材，同时，对于IT专业人员和管理人员是一个重要的资源。

第一图书网, tushu007.com

作者简介

绍洛姆·韦斯是位于纽约州约克镇的IBM预测模型小组的研究成员，同时也是美国新泽西州的罗格斯大学计算机科学专业的荣誉教授。

尼亭·因杜尔亚是澳大利亚新南威尔士大学计算机科学工程学院的讲师，同时也是数据挖掘公司Data-Miner Pty
Ltd的创始人和总裁。

张潼是美国新泽西州罗格斯大学统计与生物统计学系的教授。

第一图书网, tushu007.com

<<预测性文本挖掘基础>>

书籍目录

1
文本挖掘概述
2 从文本信息到数值向量
3 用文本进行预测
4 信息检索和文本挖掘
5
文档集的结构发现
6 在文档中查询信息
7 面向预测的数据源：数据库、混杂数据与Web
8 实例分析
9 新研究方向

第一图书网, tushu007.com

章节摘录

三致考研的同学们、朋友们亲爱的同学们、朋友们：再一次对你们辛劳备考表示由衷的慰问！
人生难得有几搏，学海无涯苦作舟，勤奋得法定成功。
我们长期辅导考研政治理论复习备考的经验证明，考生要政治理论考高分，需要有“三个掌握”：一是掌握考试大纲规定的基本理论知识的重点；二是掌握党中央的新思想、新观点、新论断以及有关的理论与实际的热点问题；三是掌握简明实用的科学方法，包括学、记、答，尤其是答题方法。
把这三者结合起来是我们师生共同的责任，我负全面的主要的责任。
一、本书的更名我主编的《任汝芬教授考研政治序列丛书》从2000年到2012年已出版13年了。
为了符合考试大纲的要求，尤其是考试的要求，使信任我们的广大考生学得好、考得好，每年我们都对本序列丛书进行了全面的深入的具体的修订与完善。
本序列丛书之三原书名为《最后冲刺》，有一些考生反映这个书名不好懂，看不出讲的什么内容，同时容易发生误解，以为这是考前的最后一本教材或参考书，同本序列丛书之四《最后四套题》出现两个“最后”，逻辑上有矛盾，忽略了点睛班及其所用教材《最后四套题》的极其重要性，故将书名更改为《形势与政策》（含高度总结、补充试题）。
本书以形势与政策冠名，有强调形势与政策的有关内容在复习考试中重要地位的深层含义，不等于是单纯的形势与政策，当然也不是有人误认为“思想政治理论考试主要是时事政治的考查”。
长期以来考研思想政治理论课考试体现了教育部根据党中央精神提高学生思想政治素质的要求，具体地说，体现了以政治性为核心的理论性、知识性、思想性、现实性、综合性的统一，引导考生自觉地拥护、贯彻以胡锦涛为总书记的党中央的理论、路线、纲领、方针、政策，成为中国特色社会主义事业合格的优秀的建设者和可靠的、优秀的接班人。
我在《序列前篇二〈课程导学〉》中曾对思想政治理论课命题或试题的主要特点有过概括：以考试大纲为范围，以权威教材为基础，以中央精神为指导，以中国特色社会主义理论体系为中心，以社会热点为理论联系实际的对象，以掌握基本理论、基本知识的重点及其运用能力为考查目标。
可以看出本课程的命题或试题既不是纯理论纯知识考查，也不是纯时事政治的考查，而是上述多方面的综合。
依据我们对多年考题的分析与综合，有2个70%：一是考查纯记忆，即“背”的分值不超过30%，而考查理解、运用分析、综合能力的占70%；二是考查基本理论、知识与中央精神、社会热点相结合的占70%。
就“形势与政策”的考查内容来看，从2006年至2009年的考试大纲规定包括：（一）中国共产党和中国政府在现阶段的重大方针政策；（二）年度间国际、国内的重大时事；（三）上述（一）、（二）项与马克思主义理论课相关知识点的适当结合。
2010年至2013年的考试大纲虽然去掉了第（三）项，但2010年至2012年的考题中仍然贯彻了这一要求。
依据上述考试情况与要求，我们在教学过程中，不仅要高度重视掌握考试大纲规定考查的基本理论、基本知识的理解记忆和运用，也要高度重视掌握形势与政策中党中央的新精神新思想、年度间的重大时事以及社会热点问题，并把它们适当地结合起来，才能确保信任我们的考生们学得好、考得好、考高分。
二、本书的依据（一）以2012年1月至10月人民日报刊登的党和政府的重大方针政策和国内外重大时事为素材。
（二）以高举中国特色社会主义的伟大旗帜为指导，深入贯彻落实科学发展观，体现党的十七大以来的新精神、新思想以及体现党的十八大新精神的胡锦涛总书记在7月23日的讲话。
有关党的十八大新精神、新思想将在十八大召开以后的冲刺班、点睛班以及《最后四套题》中充分体现。
（三）以高等教育出版社的《马克思主义理论研究和建设工程重点教材》（四本）和《考试大纲解析》为基础。
（四）以现实的理论与实际的社会热点为联系实际的对象，重点包括十一届人大有关问题、中宣部理论局编写的理论热点面对面《辩证看 务实办》、中外关系等。

第一图书网, tushu007.com

（五）以近三年考题特点为借鉴，体现试题的特点。
三、本书的任务当大家看到我们编写的政治理论复习指导序列之三的时候，正值你们的复习备考进入理论的应用阶段与总结阶段。
与这两个阶段的复习备考相配合，我们编写了复习指导序列之二、之三。
教师授课辅导的任务是以考试大纲规定的理论重点为依据，以高举中国特色社会主义伟大旗帜为统领，贯彻和体现党的十七大以来的新精神、新思想以及体现党的十八大新精神、新思想，突击时政、高度总结，预测考题、教会答题；其目标是巩固内容、增强能力、综合掌握、记忆时政。
四、本书的内容本书共五个部分，第一部分是形势与政策以及当代世界经济与政治的重点内容总结、补充试题，后四部分是分科的高度总结与补充试题。
①本书的第一部分中的形势与政策是2012年1月至2012年10月的国内外重大时事的重点内容总结和模拟试题。
这部分内容反映了党和国家现阶段的路线、方针、政策（主要是党中央的新思想、新观点、新论断）和国内外的重大时事，是形势与政策考试的现实依据和政治理论各科试题要贯彻、体现、联系的内容。
为使考生既知其然，又知其所以然，我们提供了比较完整和系统的材料。
考生对这么多的材料只需要阅读了解并对蓝色字体部分着重注意，就能理解我们编写形势与政策模拟试题及其他课程的一些模拟试题的依据和来龙去脉。
据我们和往届考生的经验，这种思路和方法事半功倍、效果显著。
我要特别提醒考生：按2013年考试大纲的规定，“形势与政策”考查年度间为2012年1月至12月，因此，我们将在序列之四《最后四套题》中对相关内容（包括形势与政策、考查内容、答题方略）精心补充完善。
②本书二至五部分编写了思想政治理论各门课程的重点内容总结和对复习指导序列之二的补充试题。
内容总结的编写思路各部分有所不同，考生一定要根据自己的习惯和特点来选择最佳的总结方式，以完成备考最后阶段的复习。
五、本书的使用我对同学们、朋友们复习备考的思路与方法建议如下：①总思路是以考试大纲规定的范围与内容为基础，以党的十七大以来的十七届三中、四中全会、五中全会、六中全会尤其党的十八大的新精神、新思想为指导，以近几年的考题为参考。
我们编写的复习指导序列之一、之二、之三都力求贯彻体现这些精神和内容。
②以做复习指导序列之二、之三的模拟练习题为中心来带动思想政治理论课的复习备考。
做题按照我们编写的教材的顺序进行。
做复习指导序列之二的每部分试题要以序列之一的相关内容为基础，并通过做题巩固序列之一的内容，掌握答题的方法，增强运用理论分析和解决问题的能力。
③做客观性试题即选择题，要精做，要问考什么、怎么考、如何答等几个为什么。
不要先看标准答案，做完一部分内容后再对答案。
这样，可以发现自己的不足和问题，然后加以解决。
对于主观性试题可以只思考回答要点和思路，然后再看参考答案，大致相近就可以了。
④需要总结和理解记忆的有三个内容：一是分部分、分层次、按顺序的各科重点理论；二是答题的思路和方法；三是形势与政策的重点试题。
关注各科的重点分析题，千万不要背题，更不能靠什么押题，只要会就行。
⑤信任我们的同学们、朋友们有两种情况：一类是参加我们辅导授课的，一类是没有参加辅导授课，只使用我们编写教材的。
对于前一类考生，听课之前一定要预习复习指导序列之二后再听课，效果较好，序列之三可以不预习。
对于后一类考生，建议按顺序反复做两遍题然后进行总结。
特别提醒：按照《要点精编》“致考研的同学们、朋友们”中的两个承诺，我已兑现了第1个承诺，9月15日后，我在人信学校等网站上作了《2013年考研思想政治考试大纲变动解读》讲座；第2个承诺在本书编写的考试大纲增加、调整的相关知识点的试题中体现。

第一图书网, tushu007.com

六、本书的后续在12月中下旬，将出版序列之四：《最后四套题》，其内容主要是我们认为是最重要的试题，以及对新信息的拾遗补阙，进一步预测考题即猜题，考生可到我们授课的考研辅导班联系获得，也可以到当地书店提前预订和购买。
我们发现多年来市场上有些资料借用了我们所编的一些模拟试题而未加说明，请考生及其他读者不要误认是我们在抄袭别人的成果。
侵权是不光彩不道德的，也是违法的。
再祝朋友们考研成功，如愿以偿！
任汝芬2012年11月5日

第一图书网, tushu007.com

<<预测性文本挖掘基础>>

编辑推荐

绍洛姆·韦斯编著的《预测性文本挖掘基础》提供了丰富的实例和软件环境，并且介绍了大量的有实际学术研究价值的内容。
本书追求实践，同时也包含了广泛的对文本挖掘有帮助的内容，既涵盖了预测学习方法，又引出了信息检索、搜索引擎和聚类等技术。
如果读者可以跟随这本书，使用我们提供的软件动手分析每一实例，相信一定可以大有收获。
本书适用于IT开发人员，管理人员，同时也适用于计算机专业的研究生。

第一图书网, tushu007.com

<<预测性文本挖掘基础>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:http://www.tushu007.com