

Amaral 期刊排名计划 JRS 及其 与 JIF 的比较分析

李 标^{1,2} 张志强¹

¹ 中国科学院国家科学图书馆兰州分馆 兰州 730000 ² 中国科学院研究生院 北京 100049

〔摘要〕 论述目前最权威的期刊评价指标——汤姆路透的期刊影响因子(JIF)的诸多缺陷及应用中的误区,介绍并分析美国西北大学学者提出的期刊排名计划(JRS),并对两者进行比较分析。提出 JRS 的科学性和可信性、快速定位高质量论文、公平性和可比较性以及稳定性等诸多优点;同时指出 JRS 缺乏动态性、灵活性、对新期刊不利以及计算繁琐等缺点。

〔关键词〕 期刊评价 期刊排名 新标准 影响因子 局限性

〔分类号〕 G350

Amaral Journal Ranking Schemes and It's Comparative Analysis with JIF

Li Li^{1,2} Zhang Zhiqiang¹

¹The Lanzhou Branch of the National Science Library of Chinese Academy of Sciences, Lanzhou 730000

²Graduate University of Chinese Academy of Sciences, Beijing 100049

〔Abstract〕 The paper thinks that although the journal impact factor(JIF) of Thomson Reuters may be the most authoritative evaluation index, there are various defects and failures in the application. The authors then introduce and analyze the new criterion system of journal ranking(JRS) proposed by the researchers of Northwestern University in America, and comparatively analyze them. The JRS has many advantages such as high credibility, comparability, stability and locating high-quality research easily, but it is not dynamic or flexible, unfair to new journal and its calculation is rather tedious.

〔Keywords〕 journal evaluating journal ranking new criterion impact index limitations

目前期刊质量评估的主要指标是期刊影响因子 JIF,该指标自身存在一些缺陷,为此,美国西北大学学者 Amaral 及其同事提出了一种新的期刊排名计划 JRS,认为其比期刊影响因子能够更加准确地评价期刊,而且可以评价特定期刊所刊载的论文,可以高效率地定位高质量的研究论文。

1 期刊影响因子及主要局限性

1.1 期刊影响因子简介

期刊影响因子是 20 世纪 60 年代由美国科学信息研究所(即现在的汤姆路透信息集团)的 Garfield 提出的,该公司出版的科学引文索引(SCI)特有的著者与著者、文献与文献之间的引用和被引用关系使影响因子的计算具有简单性和便捷性,因此使之成为目前国际上最具权威性的用于基础研究和应用基础研究科研成果的评价体系^[1]。在过去几十年中,影响因子在学术

期刊评价中发挥了重要的作用。

1.2 期刊影响因子自身局限性

期刊影响因子虽然是目前最权威的期刊评价标准,但其自身存在很多局限性:

- 源期刊库的局限。影响因子的统计局限于 SCI 源期刊库,而其存在很多缺陷。首先,该期刊库收录的期刊过少^[2],JCR 公布的 2006 版只统计了 6 166 种期刊;其次,被收录期刊的语种分布极不平衡。

- 统计过程的不规范。在影响因子的计算中,被引次数(分子)统计了刊物所有论文被引用的总次数,而刊载论文(分母)只统计论文(Articles)、简讯(Notes)和综述(Reviews)等论文类栏目的文章数^[2],对评论(Editorials)、来信(Letters)、通讯(Correspondence)和其他一些常被引用栏目的文章不予统计。事实上后者对 JIF 的贡献很大。

- 片面性。美国西北大学的 Amaral 研究小组分析了 1955 年以来 2 000 多种期刊近 2 300 万篇论文。发

现引用率最高的论文共被引用 20 万次,而其中一半论文根本没有被引用过^[3]。该研究认为平均数对论文被引率这样的非正态分布是没有充分意义的。

- 统计时间不合理。影响因子分析的是统计年之前两年刊登论文的被引用情况,但是不同学科的引用特点各不相同,比如生物、医学学科的论文引用速度较快,但是数学、经济学等领域新的研究成果甚至需要几十年来积累引用率。

- 受期刊出版周期的影响。出版周期较短的刊物更容易获得较高的影响因子,因为最先公布的成果最容易被引用。这对于出版周期较长但论文质量较高的期刊很不利。

- 著录、引用行为的不规范。论文著录格式不规范,很容易造成引文分析的偏差。作者引用中经常出现的伪引^[4]及署名中的各种不真实情况也是造成引文分析不准确的一个重要原因。

1.3 期刊影响因子使用不当的情况

- 应用范围被无限放大。影响因子评价期刊质量有一定的科学性,但利用影响因子“越俎代庖”地分析科研人员及科研机构却有待商榷。期刊的影响因子并不能反映其中每篇论文的实际被引用频次。加菲尔德也提到“用期刊的影响来评价单篇的论文或某个作者存在固有的缺陷和危险”^[5]。

- 将不可比对象利用影响因子进行比较。比如综述类期刊与原著论文期刊、不用语种的期刊、脱离同一学科领域其它刊物而单独使用某一刊物影响因子的绝对值^[6]等等。

- 根据影响因子的微小差别盲目排名期刊。影响因子的高低可从总体上反映期刊的水平,而影响因子的微小差别并没有实际意义。

2 Amaral 期刊排名计划 JRS

美国西北大学化学与生物工程学副教授 Luis Amaral 及其同事提出了一套新的期刊评价标准,该小组进行了大量的研究,并取得了大量的数据支持^[3](详见原文附录),他们认为 JRS 比影响因子能够更准确地评价学术期刊的质量和影响力。与 JIF 使用平均引用率指标不同,新体系使用数学模型来刻画期刊论文引用量的特征值,这一模型可以量化期刊论文的典型影响力及影响力范围,并提出了比较接近理想的期刊排名计划。该模型认为理想的期刊排名方法是,期刊排名越高在该刊上发现高影响力论文的可能性越高^[3]。

2.1 研究过程和结果

该研究首先获得了截止到 2006 年 12 月 Web of Science 数据库中搜索到的 22 951 535 篇论文的累积引用次数,并把限制条件定为至少 15 年内每年发表至少 50 篇论文的期刊。这一条件筛选出了发表在 2 267 种期刊上的 19 372 228 篇文章。

Amaral 提出的期刊排名计划(以下简称 JRS)中有几个引文分析重要的假设:①假设论文的质量与论文的最终影响力有着至关重要的关系,即论文被引量的近似数;②假设被引量总数和质量之间的实际联系是不确定的,也许是基于领域或基于期刊;③假设发表在期刊 J 上的论文有一个“质量”的正态分布, $q \in (\mu, \sigma)$ 。

研究的基础数据是截止到 2006 年 12 月 31 日在 Y 年发表在期刊 J 上的每篇论文的累积引用量的对数,即

$$l = \log_{10}(n), \text{ 其中, } n \text{ 为累积引用量。}$$

通过对大量期刊(详见原文附录 S2)值的概率分布 $p(l|J, Y)$ 的研究发现:①遵循近似正态的分布,每一年的分布在中心值附近呈现峰值的趋势,即有一个特征值,且稳定态阶段的平均值 $\bar{l}_{ss}(J) = \sum_{l=0}^{\infty} l' p(l'|J)$; ②每种期刊达到稳定态的时间 τ 取决于期刊本身,稳定态的函数形式为 $p_{ss}(l|J)$, 如图 1 所示。

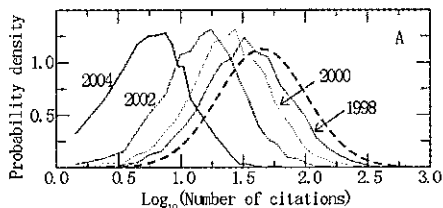


图 1 发表在给定学术期刊上的论文被引量的时间分布^[3]

上文假设③中, μ 和 σ 取决于 J。最简单的模型是把最终影响力等同于质量, $l \approx q$ 。为了使 q 的连续值映射到 n 的离散值,模型中引入另外一个参数 γ , 令

$$n = \text{floor}(10^{q-\gamma}) \quad (1)$$

可以把 γ 理解为当一篇论文被引用一次的时候 q 的取值(见图 2A)。更一般地,可以写作 $n = \text{floor}(10^{q+\varepsilon-\gamma})$ 。然而,如果 ε 不依赖 J, $\bar{l}_{ss}(J)$ 将不会明显地受 ε 影响。图 2 用散点图表示了每种期刊的经验分布及模型预测,可视化地表明该模型提供了非常接近数据的描述。

图 2 表明有一个与论文最终影响力相关的数,对

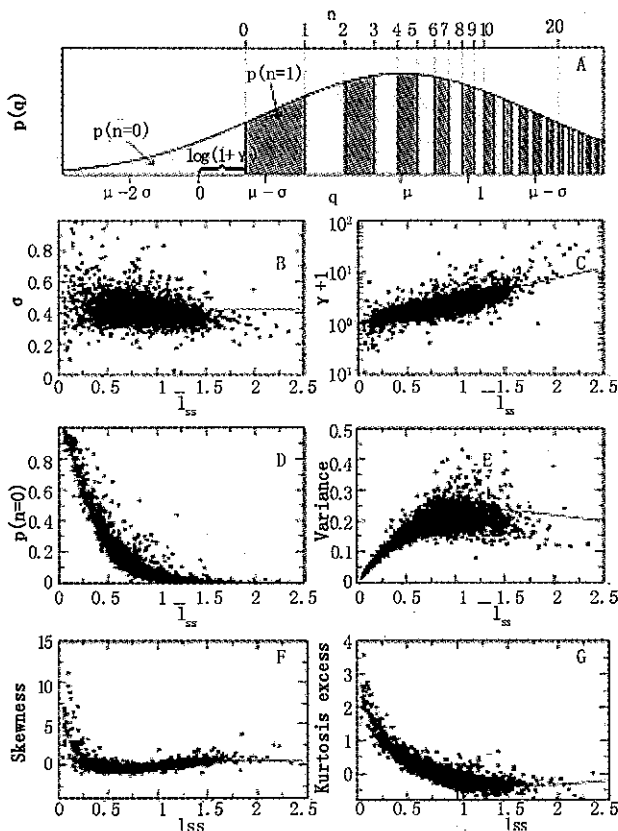


图2 给定期刊中刊载的论文被引量的稳定态分布模型^[3]

(注:图2A中, $q < \log_{10}(1+\gamma)$ 时的所有论文将不被引用, 阴影区域各不相同的面积表示了具有给定数量引文的论文的概率)

于给定期刊刊载的论文是正态分布的, 即论文质量的均值 $\bar{q}(J) = \mu$, 意味着论文质量均值 q 是该期刊发表的所有论文的 q 值的代表, 也是该期刊 q 值的代表。研究提出根据 $\bar{q}(J)$ 来排列期刊的可能。为此, 求助信息检索中的概率排序原则。这一原则规定一组期刊的理想排序是如果期刊 A 排序在 B 之前那么分别选自期刊 A 和 B 的两篇论文(a, b)的 $q(a) > q(b)$ 的概率最大。概率也就是多级 AUC 统。

研究发现根据 $\bar{q}(J)$ 来排序期刊具有比 JIF 更高的准确性^[3]。该研究论文及附件 S3 选取 ISI 不同学科的大量期刊, 分别根据 $\bar{q}(J)$ 和 JIF 排名后与 AUC 理想排名进行对比, 并对结果做了可视化显示。全部的比较结果都显示根据 $\bar{q}(J)$ 排名期刊比 JIF 排名准确得多。

2.2 研究方法

2.2.1 识别稳定态 用 $p(1|J, Y)$ 的时间演变来识别稳定态和非稳定态阶段。其中稳定态 $\frac{d\bar{l}(J, Y)}{dY} \approx 0$ 。定义非稳定态时期 $\tau = 2006 - \gamma_0$, 其中 γ_0 为移动平均数 < 0.005 时 Y 的最大值。

为了确定引文分布稳定的状态, 该研究小组用 Kolmogorov-Smirnov 检验来比较多组年代的引文分布, 并调整 P 值矩阵合适的正方形对角线模型。他们发现该方法可以确定引文分布达到稳定的阶段。研究用来比较的分布是 Y0 之前最近的稳定时期。

2.2.2 估计期刊的 μ, σ 和 γ 对于每个稳定态的引文分布而言, 该模型(见等式(1))有三个参数需要估计: μ, σ 和 γ 。该小组研究发现不存在这一模型参数的最大可能性估计, 因此通过最大化 X^2 统计量来估计参数(参见原文附录 S4)

$$X^2 = \sum_n \frac{[p_n - p(n|\mu, \sigma, \gamma)]^2}{p(n|\mu, \sigma, \gamma)} \quad (2)$$

其中 p_n 为被引量为 n 的论文, $p(n|\mu, \sigma, \gamma)$ 为模型中存在被引量为 n 的论文的概率。

$$p(n|\mu, \sigma, \gamma) = \begin{cases} \int_{-\infty}^{\log_{10}(\gamma+1)} \frac{dq}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(q-\mu)^2}{2\sigma^2}\right] & n=0 \\ \int_{\log_{10}(n+\gamma)}^{\log_{10}(n+\gamma+1)} \frac{dq}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(q-\mu)^2}{2\sigma^2}\right] & n \geq 1 \end{cases} \quad (3)$$

研究过程中研究人员分开存储数据以使每个样本中有 10 个数据。那么 X^2 为

$$\frac{[\hat{p}_{n_1, n_2} - \hat{p}([n_1, n_2]|\mu, \sigma, \gamma)]^2}{\hat{p}([n_1, n_2]|\mu, \sigma, \gamma)} \quad (4)$$

其中, $\hat{p}_{n_1, n_2} = \sum_{n_1 \leq n_2} \frac{P_n}{n_2 - n_1 + 1}$, $\hat{p}([n_1, n_2]|\mu, \sigma, \gamma) = \sum_{n_1 \leq n_2} p(n|\mu, \sigma, \gamma)$ 。合适的参数意味着 σ 弱依赖于 \bar{l} (见图 2B)。相反, γ 强依赖于 \bar{l} (见图 2C)

$$\gamma(\bar{l}) + 1 = C_0 e^{C_1 \bar{l}} \quad (5)$$

设 $C_0 = 0.91 + 0.02$, $C_1 = 1.03 + 0.02$ 。简单来说, 当比较实验分布与模型预测(见图 2D-G)时, 假设 $\sigma = 0.419$, $\gamma \bar{l} = -1 + \exp(\bar{l})$ 。假设这两个依赖, 得到 μ 和 \bar{l} 的关系为

$$\bar{l} = \frac{\sum_{n=1}^{\infty} p(n|\mu, \sigma, \gamma(\bar{l})) \log_{10}(n)}{1 - p(0|\mu, \sigma, \gamma(\bar{l}))} \quad (6)$$

正如图 2C 中所示, γ 的估计值显示了很大的波动, 因此通过最小化 X^2 来估计 μ 和 σ 。该研究小组证明 $\mu = \bar{q}$ 是我们用来排名期刊的最好方法^[3]。

2.2.3 计算多级 AUC 该研究把最好的排序定义为使多级 AUC 统计量取得最大值。对于一组期刊 $F = \{J_x\}$ 和期刊排名 R, 我们定义多级 AUC 统计 $M(F, R)$ 为

$$M(F, R) = \sum_{\substack{(J_A, J_B) \in F \\ R(A) < R(B)}} W_{AB} P_{AB}(R) \quad (7)$$

用 P_{AB} 表示分别选自期刊 J_A, J_B 的一组给定文章(a, b), 且 $q(a) > q(b)$ 。用 W_{AB} 表示我们赋给每个概率的权重, 这依赖于在稳定态阶段发表在期刊 J_A, J_B 上的论文数 N_A 和 N_B , 如下

$$w_{AB} = \frac{N_A N_B}{\sum_{i \in j} N_i N_j} \quad (8)$$

原则上,可以计算期刊被引量分布的每种排列的多级 AUC 统计量,并选择取得最高值的排序。然而所有排列的最保守数目都大得难以处理,幸好几乎所有案例都遵从传递性的特征,即如果 $a > b$ 且 $b > c$, 则 $a > c$, 这简化了最优化的任务。在很少的例子中传递性不适用,可以求助于 brute-force 优化,通过改变每个分布的顺序和求出多级 AUC 统计量最大值的方法来解除排序中的不明确性^[3]。

3 JRS 与 JIF 的比较分析

3.1 JRS 的“比较优势”

JRS 相对于期刊影响因子有很多改进,其自身的方法体系解决了影响因子的几个主要缺陷,具有科学性。与 JIF 相比,JRS 的主要优势有:

3.1.1 具有较高的科学性和可信性 JRS 采用每篇论文被引量的对数作为研究的基础数据。对不同学科领域、不同“年龄”期刊的研究发现,1 的分布都近似符合正态分布,以 1 为研究对象巧妙地避免了同一期刊中的论文被引量相差悬殊的情况,这样可以通过更加接近的方法减弱少数高被引论文对评价结果的干扰,从而使该评价方法具有充分的意义。而影响因子的研究对象是某期刊刊载的所有论文的被引频次,期刊所有论文被引频次并不符合正态分布,对于非正态分布的研究对象而言,在宽泛的研究范围内取平均值是没有充分意义的。

3.1.2 便于快速锁定高质量论文 JRS 以每篇论文被引量的对数为基础数据,而研究发现该数据近似服从正态分布,进一步假设论文的质量符合正态分布,这样巧妙地将论文质量与论文被引量的对数联系起来,每篇论文的质量可以很方便地得到估价。同时,JRS 的研究给出了预测论文取得特征值的速度,这样可以更加科学地对期刊论文进行评价,最大程度地避免了学科差别引起的被引规律的不同。JRS 提出理想排名计划,通过 AUC 统计进行排名,基本可以做到在排名较高的期刊中发现高质量论文的几率最高,因此可以更加方便、准确地定位高质量的研究论文。而使用影响因子更多的是“以刊取文”,通过期刊影响因子的高低来衡量论文的质量,这具有很高的盲目性和不准确性。

3.1.3 更具有公平性和可比较性 选择“稳定态”阶段进行研究,可以最大限度地避免学科领域的差别对

期刊质量评价的影响,而且稳定态阶段的计量可以排除很多不稳定因素的影响。因为 Amaral 小组研究发现,很多学科的研究论文甚至需要 26 年才能完成被引量的累积,而在稳定态期刊刊载的论文被引量的对数趋于稳定。该小组还提出可以利用数学模型来预测不同期刊论文取得“特征值”的速度。而影响因子的统计时间始终是统计年之前两年,很容易造成那些需要很长时间来累计被引量的学科被低估。

3.1.4 具有较高的稳定性 JRS 是一个长期指标,短期内不易受到人工干预而出现较大幅度的波动情况。JRS 是基于论文被引量的对数符合正态分布这一实验发现的,通过“稳定态”各项指标的计算避免了期刊质量的评价出现较大的偏差。如上所述,期刊影响因子分析的是统计年之前两年期刊论文被引用情况,是一个短期的指标,很容易受到人工干预,造成评价的不稳定性。现实中通过自引来提高影响因子的情况不胜枚举,这种现象的出现与影响因子指标本身的不足有很大关系。

3.2 JRS 的“比较劣势”

3.2.1 缺乏动态性 期刊的“稳定态”是一段孤立、静止存在的阶段,对该阶段的分析得到的是期刊的一种类似 h 指数的“终身成就”,容易造成评估结果的固定,不能及时地反映非稳定态阶段期刊乃至论文的质量和影响力。而期刊影响因子是一个动态指标,随时间推移该指标是动态变化的,充分反映了该评价标准的时间敏感性。

3.2.2 对新期刊非常不利 JRS 是一种累积指标,对新期刊非常不利。因为要达到 JRS 中要求的“稳定态”阶段需要一段时间,该研究小组也提出,某些学科的论文甚至需要 26 年才能完成其累积被引,亦即不少期刊需要一个相当长的阶段才能达到“稳定态”,这会造成尚未达到稳定态的新期刊在较长时间内其影响力无法统计的尴尬状况。

3.2.3 缺乏灵活性 JRS 对期刊近期的被引用情况不予关注,不能及时反映期刊的动态,这样不利于期刊编辑及时发现期刊潜在的问题和面对的竞争,不利于形成鼓励创新和发展的氛围,很容易造成一些指标比较高的期刊“吃老本”的情况,导致一些期刊的质量停滞不前甚至倒退。

3.2.4 统计计算繁琐 与影响因子、h 指数不同的是,JRS 完全颠覆了以往评价方法的简单便捷的计算方式,代之以复杂的数学模型,这会给期刊的评价分析工作带来不小的挑战。

4 结 论

目前很多国家在进行科学评价时过于依赖期刊影响因子,但影响因子具有很多局限性,且存在很多不恰当应用的现象。美国西北大学学者提出的期刊评价JRS针对影响因子的缺陷进行了修正。首先,JRS使用近似正态分布的论文被引量对数作为研究对象,更具有科学性和可信性。其次,JRS可以预测不同领域期刊论文获得引用“特征值”的速度,为评价特定期刊提供了较为固定的模式。再次,该标准弥补了影响因子无法评价论文的缺憾。另外,JRS具有较好的稳定性。虽然有如此种种优点,JRS本身仍然存在一些不足之处,需要国内外学者的讨论和完善。

参考文献:

[1] 靳小青. SCI期刊的影响因子剖析. 情报科学, 2001(10):1054

[作者简介] 李 标,女,1982年生,硕士研究生。

张志强,男,1964年生,研究员,馆长,博士生导师,发表论文150余篇,出版专著和编著5部、译著5部。

(上接第142页)

访问、用户统一管理和认证授权等。组合式 Portal 将 Web 服务通过 Portal 即插即用,使其更加灵活。

5 结 语

虚拟企业信息系统体系是虚拟企业信息系统的软件部件及其相互之间的关系,要求其具有良好的可集成性、可重构性、流程管理及相应的安全性,但虚拟企业临时性和动态性的特点给虚拟企业的信息集成带来了挑战。本文从 SOA 和 Web 服务技术的角度研究分析了虚拟企业信息集成的模型及实现流程,通过 Web 服务技术的动态服务绑定解决虚拟企业的动态性和临时性问题。基于 Web 服务的 SOA 虚拟企业信息集成方案,可以方便地实现已有系统、新开发的 Web 服务应用等各种系统的集成。这种松散的联合和动态的集成,对建立无缝跨平台互操作的信息共享与数据交换具有极大的优势。基于 Web 服务的开放式企业信息集成模型,既可以满足企业信息集成的需要,又因为引入了对 Web 服务和 SOA 的支持,能够最大限度地降低企业信息集成的复杂度,提高整个系统的灵活性。

[作者简介] 余望枝,女,1974年生,馆员,发表论文9篇,编著1部。

邓 武,男,1976年生,讲师,硕士,发表论文20余篇,参编专著1部。

-1056.

- [2] 任胜利,王宝庆. 应慎重使用期刊的影响因子评价科研成果. 科学通报, 2000(2):218-222.
- [3] Stringer M J, Marca S P, Nunes Amaral L A. Effectiveness of journal ranking schemes as a tool for locating information. PLoS ONE, 2008, 3(2). [2008-03-10]. <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0001683>.
- [4] 李菡,刘振民,刘改换. 科技期刊影响因子与科研绩效评价. 太原理工大学学报:社会科学版, 2006(4):76-78.
- [5] Garfield E. The History and Meaning of the Journal Impact Factor. JAMA - Journal Of The American Medical Association, 2006, 295(1):90-93.
- [6] 吴琦. 科技期刊的影响因子正确使用和误用. 中国基础科学, 2001(7):29-34.
- [7] 任霄鹏. 不用影响因子? 科学家提出期刊评价新标准. [2008-03-10]. <http://www.sciencenet.cn/htmlnews/20083615502548202900.html>.

参考文献:

- [1] Erasala N, David C. Enterprise Application Integration in the electronic business world, Computer Standards & Interfaces, 2003, 25(2):69-82.
- [2] Umar A. The emerging role of the web for enterprise applications and ASPs//Proceedings of the IEEE on Evolution of Internet Technologies, 2004, 92(9):1420-1438.
- [3] Bertino E, Martino L. Security in SOA and Web services//2006 IEEE International Conference on Services Computing, SCC 2006, Chicago, IL, Institute of Electrical and Electronics Engineers Computer Society, Piscataway, xli.
- [4] Epstein J, Matsumoto S, McGraw G. Software security and SOA: Danger, will robinson. IEEE Security and Privacy, 2006, 4(1):80-83.
- [5] Howerton J T. Services-oriented architecture and Web 2.0. IT Professional, 2007, 9(3):62-64.
- [6] Stal M. Using architectural patterns and blueprints for services-oriented architecture. IEEE Software, 2006, 23(2):54-61.
- [7] Luo M, Goldshlager B, Zhang L J. Designing and implementing Enterprise Services Bus (ESB) and SOA solutions//IEEE International Conference on Services Computing, ICWS 2005, Orlando, FL, of Electrical and Electronics Engineers Computer Society, Piscataway, xiv.