

特定领域问答网站中的标签自然顺序研究

贾俊芳¹, 李国强²

¹山西大同大学计算机与网络工程学院, 中国大同市, 037009

²上海交通大学软件学院, 中国上海市, 200240

摘要: 标注是Web 2.0的一个重要特征。它使得社会计算系统（如问答网站）的用户们可以自由地标记内容。然而，标注真的是自由不受限的吗？现有工作表明，用户们常常可以隐性地就哪种标签最能描述在线社区的内容达成共识。然而，目前还没有针对用户在标注过程中对标签排序的规律性开展研究。本文专注于研究特定领域问答网站中的标签自然排序，并对CodeProject, SegmentFault, Biostars以及CareerCup 4个问答网站上数以百万计的问题中的标签序列进行研究。结果表明，这些问答网站的用户可以就问题标签的排序达成隐性共识。研究了标签之间的关系，这些关系可以解释标签自然顺序的出现。该研究为利用标签的自然顺序提升现有标签推荐以及问答站点导航提供了可能。

关键词: 问答网站；标注；自然顺序；Skip gram

<https://doi.org/10.1631/FITEE.1900645>