

推进规范文档建设

规范文档（Authority File）是指由规范记录组成的计算机文档。通过规范文档控制文档标目的前后一致和唯一性，可以对书目记录的检索点进行规范化校验，提升检索效率；对学术实体名称进行归一化处理，为开展实体计量、分析、关联以及评价奠定基础。

自20世纪初期西方图书馆学家卡特提出规范控制概念以来，各国一方面强化了规范控制的标准制订工作，另一方面推进了对各类实体名称规范文档的建设工作。1977年美国国会图书馆推动“名称规范合作计划”，2001年欧洲的15家机构合作开展“国际规范资料库合作建制计划”，2003年德国国家图书馆、美国国会图书馆和OCLC开发“虚拟国际规范文档”。1995年我国国家图书馆开始名称规范工作，涉及个人名称、团体名称、统一题名及丛编题名，截止2013年，其规范数据库收录123万多条记录。2000年由CALIS牵头，国家图书馆、香港地区大学图书馆和台湾汉学研究中心合作建设中文名称规范数据库，截止2013年收录88万条记录。

规范文档建设需要围绕一个特定名称实体的所有数据资源集中表示，确立名称实体的语义关系，并运用形式化语言表述，支持数据的语义搜索和导航，实现数据聚合，发挥其在网络环境中的规范控制作用。这里需要通过各种算法和人工判断进行实体识别，通过规范控制模型及规则分析来实现实体名称的归一，并建立实体间语义描述，当然，规范文档的维护更新需要推进协同应用，才能共建共享和持续发展。

随着海量数据的不断产生，资源组织和数据处理需要大力加强各类实体名称规范文档的构建，提升对实体识别的检索效率，强化实体归一的规范控制，促进实体计算的完整聚合，挖掘实体特征的语义关系，推进实体名称语义层面的协同应用，发挥规范文档作为重要知识组织工具在语义网环境中的作用。规范文档建设和应用既可以深化规范控制原理，又利于丰富知识组织体系。

■ 曾建勋