



Evento	Salão UFRGS 2014: SIC - XXVI SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2014
Local	Porto Alegre
Título	Produção científica em colaboração entre Espanha e Brasil: metodologia para normalização de nomes institucionais
Autor	ALESSANDRA DE LACERDA SCATIGNO
Orientador	SONIA ELISA CAREGNATO

Objetiva explicar a metodologia adotada na fase de normalização de dados institucionais dentro de pesquisa realizada pelo Grupo de Pesquisa em Comunicação Científica da FABICO/UFRGS em conjunto com a Universidade Carlos III de Madrid (UC3M). Intitulada “Estudo da produção científica em colaboração entre Espanha e Brasil: características e impacto da produção conjunta”, a pesquisa, ainda em realização, visa conhecer as relações de cooperação científica estabelecidas entre os dois países por meio de uso de indicadores bibliométricos. Para isso, coletaram-se os registros relativos à produção de ambos os países no período 2006-2011 na base de dados *Web of Science* (WoS) da Thomson-Reuters. Foram descarregados todos os 620.252 registros que continham endereços de pelo menos uma instituição brasileira ou espanhola. Porém, observou-se que os nomes das instituições apareciam redigidos de diferentes formas, gerando resultados pouco confiáveis. A fim de padronizar os nomes das instituições, o Laboratório de Estudos Métricos da Informação (LEMI) da UC3M desenvolveu uma ferramenta chamada “Normalizador”, que opera em uma plataforma *web*, recolhendo todos os endereços institucionais dos registros descarregados. Nesta etapa, os endereços já receberam uma atribuição automática do país ao qual pertencem, mas ainda necessitam ser agrupados por similaridade. Primeiramente, para cada instituição, foram criadas “regras” que dispõem das mais variadas formas de como ela pode ser denominada. Para encontrar o maior número possível de variações, foram realizadas buscas na internet. Após a conferência de que as regras recuperavam somente endereços da instituição em questão, aplicou-se a “etiqueta”, que contém um conjunto de regras. Posteriormente, fez-se uma lista de endereços que ainda não tinham sido etiquetados e, a partir deles, criou-se novas regras e etiquetas com o intuito de padronizar o maior número possível de endereços. Até o momento, foram normalizadas 176 instituições brasileiras, porém este número tende a aumentar a medida que o trabalho é realizado. Por fim, cada etiqueta representará uma instituição e, assim, será possível obter os principais indicadores bibliométricos. Com a obtenção destes, se consegue identificar as características da colaboração entre os dois países. E, logo, fazer uma análise aprofundada das áreas nas quais a colaboração é mais acentuada, assim como observar quais são as instituições participantes nas áreas de maior produção.