
A visibilidade no Google Scholar dos repositórios digitais de acesso aberto brasileiros e portugueses

Michelli Costa

Universidade de Brasília

michellicosta11@gmail.com

Resumo

A visibilidade dos repositórios em motores de busca web é fundamental para que eles possam, de fato, cumprir com sua função de ampliar a disseminação da produção acadêmica na Internet. No entanto, estudos realizados nos últimos anos apontam a baixa indexação dos documentos depositados em repositórios nos motores de busca, em especial o Google Scholar. Partindo de metodologias já utilizadas anteriormente, este estudo analisou a taxa de indexação no Google Scholar dos repositórios brasileiros e portugueses listados pelo Ranking Web of Repositories. Como resultado, foi identificado que os repositórios analisados apresentaram taxas significativamente melhores do que as encontradas por outros estudos, mas ainda não são satisfatórias. Além disto, foi observada a tendência à concentração dos documentos em pequenos grupos de repositórios. De forma geral, é possível concluir que os repositórios brasileiros e portugueses possuem mais visibilidade no Google Scholar do que o esperado, mas ainda aquém de suas possibilidades.

Palavras-chave: Repositórios digitais. Acesso Aberto. Indexação. Google Scholar. Visibilidade web.

Abstract

The repositories visibility on web search engines is essential to implement their function to expand the dissemination of academic output on the Internet. However, studies conducted in recent years indicate low indexing in search engines of documents from repositories, especially in Google Scholar. Using methodologies already known, this research examined in Google Scholar the indexation rates from Brazilian and Portuguese repositories listed by Ranking Web of Repositories. As result it was identified that repositories analyzed showed significantly better rates than those rates found by other studies, but they are still not satisfactory. Moreover, a trend was observed in concentration of documents in small groups of repositories. This study concluded that both, Brazilians and Portuguese repositories, are visible in Google Scholar, but their results may be better than presented now.

Key-words: Digital repositories. Open Access. Indexing. Google Scholar. Web visibility.

Introdução

A Budapest Open Access Initiative (BOAI, 2002) propõe como uma das estratégias para a promoção do acesso aberto à informação científica o depósito, pelos próprios acadêmicos, dos artigos publicados por revistas arbitradas, em arquivos abertos e eletrônicos, que mais tarde foram nomeados repositórios digitais. Ainda segundo a BOAI (2002), os repositórios devem estar em conformidade com as recomendações Open Archives Initiative (OAI), de modo que possam ter seu conteúdo localizado por motores de busca na Internet. As orientações da BOAI (2002) são fundamentais para o cumprimento de uma das finalidades dos repositórios, que é tornar a produção acadêmica de uma instituição amplamente visível, conforme destaca Russell (2011). Diante da importância da visibilidade dos repositórios na Internet, Arlitsch e O'Brien (2012) realizaram pesquisa a partir do Search Engine Optimization (SEO) e identificaram que os repositórios dos Estados Unidos da América possuíam baixas taxas de indexação no Google Scholar.

Orduña-Malea e López-Cózar (2014) também afirmam que os repositórios são um dos componentes-chave do acesso aberto. Para sustentar a afirmativa os autores citam o estudo de Archambault, no qual foi identificado que mais de 75% dos artigos disponíveis em acesso aberto na Internet estão nessa condição por meio de repositórios e não pelas próprias revistas. Além disso, destacam a importância da visibilidade dos repositórios na web, pois isto “é essencial para garantir que o conteúdo que eles armazenam permita o acesso aberto efetivamente verdadeiro para a comunidade” (p.2, tradução nossa). Diante da questão, Orduña-Malea e López-Cózar (2014) realizaram um estudo para determinar a visibilidade e o impacto no Google e no Google Scholar dos repositórios institucionais da América Latina listados no Ranking Web of Repositories. Como um dos resultados, identificaram baixa quantidade de arquivos PDF coletados pelo Google Scholar, corroborando assim com os resultados do estudo de Arlitsch e O'Brien.

Conforme fora apontado por López-Cózar e Robinson-Garcia (2012), o Ranking Web of Repositories é o instrumento de avaliação de impacto na web com maior alcance para o contexto dos repositórios. Segundo Aguillo et al (2010), além do impacto, o Ranking também avalia, na categoria visibilidade, a indexação de documentos por motores de busca e a quantidade de links externos que os repositórios recebem no Yahoo e no Exalead. Portanto, ele é uma importante referência para as discussões sobre o tema e um valioso ponto de partida para novos estudos.

Considerando o contexto de discussão atual sobre a visibilidade dos repositórios na Internet, em especial no motor de busca Google Scholar, este estudo buscou responder a

seguinte questão: *Qual a taxa de visibilidade no Google Scholar dos repositórios brasileiros e portugueses listados pelo Ranking Web of Repositories?*

Metodologia

Para responder a pergunta do estudo foi adotada, como metodologia, a pesquisa documental e a análise estatística dos resultados. Os procedimentos metodológicos para tanto foram sistematizados em cinco passos.

O primeiro passo foi a seleção do universo de análise, que utilizou como fonte de informação o Ranking Web of Repositories. A partir da navegação pelas listas dos países, foram identificados todos os sistemas descritos nas páginas “Brasil” e “Portugal”.

Passo	Objetivo	Fonte de informações	Método de coleta
1.	Identificar os repositórios a serem analisados	Lista do Ranking Web of Repositories	Pesquisa documental (navegação pelas listas dos países)
2.	Identificar a quantidade total de documentos nos repositórios selecionados	Informações disponibilizadas pelos repositórios (lista de títulos e informações estatísticas)	Pesquisa documental (navegação pelos índices de títulos e páginas sobre as estatísticas)
3.	Identificar a quantidade de documentos recuperados no Google Scholar de cada repositório selecionado	Informações disponibilizadas pelo Google Scholar	Pesquisa documental (pesquisa no campo de busca da página inicial com o seguinte argumento site:urldorepositório)
4.	Calcular a taxa de visibilidade dos repositórios selecionados	Informações coletadas nos repositórios selecionados (Doc.RD) e os resultados das buscas realizadas no Google Scholar (Doc.GS)	Apliação da fórmula TI = Doc.GS*100/Doc.RD
5.	Analisar os resultados obtidos	As informações coletadas nos repositórios selecionados (Doc.RD), os resultados das buscas realizadas no Google Scholar (Doc.GS), a taxa de visibilidade (TI) e a posição dos repositórios selecionados no Ranking Web of Repositories	Análise estatística dos resultados.

Quadro 1: Quadro resumo da metodologia

Após a identificação, foi levantada a quantidade total de documentos disponíveis em cada repositório. As informações foram coletadas nos índices de “Títulos” disponíveis nos sistemas, quando aplicável. Nos sistemas que não dispunham desta funcionalidade foram pesquisadas as estatísticas sobre a quantidade de documentos informada pelos próprios sistemas.

Paralelamente à identificação da quantidade total dos documentos nos repositórios, foi levantada a quantidade de documentos recuperados na busca do Google Scholar. Neste passo, foi utilizado o campo de busca simples, na página inicial do site. Para a pesquisa foi utilizado o argumento de busca [site:urldorepositório], mesmo método utilizado por Arlitsch

e O'Brien (2012). Como resultado da busca, foram consideradas as informações geradas pelo próprio buscador.

De posse das duas informações, levantadas nos passos dois e três, foi calculada a taxa de visibilidade (TV) dos repositórios selecionados no Google Scholar. Para cada repositório foi aplicada a seguinte fórmula [TV = quant. de doc. no Google Scholar * 100 / quant. de documentos totais do repositório].

A partir dos resultados obtidos nos passos dois, três e quatro, foram elaborados rankings sobre os três aspectos analisados. Somados a eles, foi ainda adicionada a posição dos repositórios no Ranking Web of Repositories. O conjunto destes resultados permitiu uma análise descritiva e estatística dos resultados.

Repositórios digitais brasileiros e portugueses

A amostra dos repositórios digitais brasileiros e portugueses foi definida a partir das listas geradas pelo Ranking Web of Repositories. Delas foram identificados todos os sistemas descritos nos dois países. O uso do Ranking como fonte de informação para esta pesquisa justifica-se pelo seu universo de análise e por seus indicadores de avaliação. Segundo Aguillo et al (2010), para a definição do universo de análise são realizados levantamentos nos dois principais diretórios internacionais de repositórios: Directory of Open Access Repositories (OpenDOAR) e Registry for Open Access Repositories (ROAR). Do conjunto inicial, identificado pelo Ranking, são excluídos os repositórios sem conteúdo acadêmico, portais de periódicos e repositórios que não possuem domínio ou subdomínios próprios. A outra característica do Ranking, que justifica sua seleção, é o uso de indicadores gerados a partir de motores de busca para sua avaliação. Ainda de acordo com Aguillo et al (2010), são utilizadas quatro categorias para descrever a presença e o impacto dos repositórios na web. Portanto, o Ranking reúne um conjunto significativo dos repositórios com publicações acadêmicas dos países e avalia aspectos sobre a visibilidade na web desses sistemas.

Na lista dos repositórios brasileiros foram apresentados 45 sistemas, mas três foram excluídos da análise. Na lista portuguesa identificou-se 41 repositórios e quatro foram excluídos. Três exclusões foram decorrentes da falta de acesso à página web dos repositórios e três exclusões aconteceram devido à natureza dos sistemas, que são agregadores do conteúdo. Por fim, um último sistema foi excluído por ter sido apresentado repetidamente. Os dados apresentados foram levantados na primeira quinzena de setembro.

Os repositórios selecionados foram brevemente caracterizados segundo a natureza dos seus documentos e o software utilizado. Com isto foram utilizadas três categorias de classificação. A primeira delas foi "repositório institucional" para os sistemas que apresentavam exclusivamente a produção científica de sua instituição. A segunda foi "repositório digital" para os sistemas que armazenavam, além da produção científica,

documentação administrativas e de outras naturezas não-acadêmicas. Por fim, utilizou-se a classificação “repositório de teses e dissertações” para os sistemas que se limitavam ao tratamento deste tipo de documento.

A partir das categorias descritas foi identificado que no Brasil os repositórios institucionais representam 53% (22 de 42). A outra metade está representada pelos repositórios digitais 33% (14 de 42) e 14% (6 de 42) de repositórios de teses e dissertações. Já entre os portugueses foi observada uma quantidade significativamente superior na porcentagem de repositórios institucionais, esta proporção está para 89% (33 de 37). No universo português foram localizados 4 “repositórios digitais”, o que representa apenas 11% do universo.

Entre os repositórios brasileiros observou-se que existe uma variedade maior entre as naturezas dos repositórios, uma vez que cerca de metade deles são destinados para o tratamento de teses e dissertações ou produções acadêmicas e administrativas de uma instituição. Já entre os repositórios portugueses nota-se a maior concentração no tratamento da produção acadêmica com a presença, necessariamente, dos artigos de periódicos revisados por partes.

Também foi identificado que os repositórios portugueses utilizam majoritariamente (97%) o software DSpace. Apenas um repositório, dentre os analisados utiliza o software Digitool. Entre os brasileiros o uso do DSpace também é majoritário, mas em menores proporções (76%). No país, além do DSpace foram identificados repositórios em plataformas TEDE (7%) e Nou-Raw (12%), que são tecnologias nacionais. Além destes, foram identificados dois sistemas (5%) cujos softwares não foram classificados em nenhum grupo citado. De forma geral, foi observado o uso majoritário do software DSpace entre os repositórios que tratam exclusivamente da produção acadêmica da instituição, com a presença de artigos de periódicos.

Indexação dos repositórios brasileiros e portugueses

A taxa de indexação proposta foi calculada a partir da porcentagem dos documentos dos repositórios que são recuperados no Google Scholar. Portanto, foi necessária a identificação de duas informações sobre cada repositório: a quantidade de documentos presentes nos sistemas e a quantidade de documentos recuperados pelo Google Scholar. Para auxiliar na análise dos resultados foram elaborados rankings sobre os valores referentes à quantidade total de documentos, quantidade de documentos recuperados pelo Google Scholar e a taxa de visibilidade (Tabela 1 e Tabela 2).

Quantidade de documentos disponíveis nos repositórios

Dentre os 42 repositórios brasileiros foram identificados aproximadamente 604 mil documentos (docs). No entanto, observou-se que a quantidade de documentos por

repositório se dá de forma irregular, pois, cinco repositórios concentram pouco mais da metade (54%) do total de documentos do país. Os cinco repositórios com maior volume de documentos são:

1. Repositório Digital Universidade Federal do Rio Grande do Sul LUME – 99.632 docs.;
2. Repositório Institucional UNESP Universidade Estadual Paulista Júlio de Mesquita Filho – 69.382 docs;
3. Biblioteca Digital da UNICAMP Universidade de Campinas – 56.257 docs;
4. Alice Repository Open Access to Scientific Information Embrapa – 51.981 docs;
5. Repositório Institucional Universidade de Santa Catarina – 50.340 docs.

Já entre os repositórios portugueses foram identificados aproximadamente 212 mil documentos nos 37 repositórios. O mesmo padrão na distribuição dos documentos observada nos repositórios brasileiros foi encontrado entre os portugueses. Cinco repositórios concentram quase metade (49%) do total de documentos do país, são eles:

1. Repositório Aberto Universidade do Porto – 33.504 docs;
2. Repositório da Universidade do Minho – 28.201 docs;
3. Estudo Geral Repositório Digital da Universidade de Coimbra – 16.945 docs;
4. Repositório Institucional Universidade Católica Portuguesa Veritati – 12.372 docs;
5. Repositorio Biblioteca da Faculdade de Engenharia da Universidade do Porto – 12.177 docs.

Quantidade de documentos recuperados pelo Google Scholar

No Brasil, em termos de documentos recuperados pelo Google Scholar, três repositórios se mantiveram no grupo dos cinco com maior quantidade de documentos. No entanto, manteve-se a tendência de cinco repositórios do conjunto concentrarem a metade, ou mais da metade, de todos os documentos recuperados. No total, foram recuperados no Google Scholar mais de 282 mil documentos, destes 77% são providos pelos seguintes sistemas:

1. Repositório Digital Universidade Federal do Rio Grande do Sul LUME – 75.600 docs;
2. Universidade de São Paulo Biblioteca Digital de Teses e Dissertações – 44.000 docs;
3. Repositório Institucional Universidade Federal de Santa Catarina – 43.600 docs;
4. Alice Repository Open Access to Scientific Information Embrapa – 30.100 docs;
5. Biblioteca Digital da UNICAMP Universidade de Campinas – 22.500 docs;

Novamente, o mesmo padrão encontrado entre os repositórios brasileiros foi observado entre os repositórios portugueses. Apenas três repositórios daqueles com maior quantidade de documentos depositados mantiveram-se entre os cinco primeiros quando foi analisada a quantidade de documentos recuperados pelo Google Scholar. Além disso, os cinco repositórios com maior quantidade de documentos recuperados reuniram mais da metade (54%) dos cerca de 145 mil documentos recuperados pelo Google Scholar nos repositórios portugueses. Os cinco repositórios que se destacaram neste grupo foram:

1. Repositório Aberto Universidade do Porto – 28.100 docs;
2. Universidade do Minho Repositorium – 20.400 docs;
3. Repositório Institucional Universidade Católica Portuguesa Veritati – 10.200 docs;
4. Repositório Institucional da Universidade de Aveiro RIA – 9.390 docs;
5. Repositório Universidade de Lisboa – 9.190 docs.

Taxa de indexação dos repositórios no Google Scholar

De posse das duas informações – quantidade total dos documentos no repositório e quantidade de documentos recuperados pelo Google Scholar – foi possível calcular a porcentagem dos documentos dos repositórios que são recuperados pelo Google Scholar, que neste estudo foi denominado de taxa de indexação – TI (Figura 1).

$$TI = \frac{\text{quant. de docs. no Google Scholar} \times 100}{\text{quant. de documentos totais no repositório}}$$

Figura 1: Fórmula para calcular a taxa de indexação

Como resultado observou-se que os repositórios com maiores taxas de indexação não coincidiu com aqueles com maiores quantidades de documentos totais, tampouco com aqueles que possuem mais resultados de busca no Google Scholar. O fenômeno foi observado tanto nos repositórios brasileiros quanto nos portugueses.

No Brasil, os repositórios com maiores taxas de visibilidade são:

1. Repositório Institucional Universidade Federal de Goiás – 274,7%;
2. Repositório Institucional Centro Universitário de Brasília – 99,7%;
3. Repositório de Outras Coleções Abertas Universidade Tecnológica Federal do Paraná ROCA – 99,6%;
4. Repositório Digital da Universidade Federal do Recôncavo da Bahia – 98,1%;
5. Repositorio Universidade de Caxias do Sul – 97,1%.

Já entre os repositórios portugueses destacaram-se os seguintes sistemas:

1. REPAP Repositório de Administração Pública –100,7%;
2. Repositório Científico Digital da Universidade da Madeira – 94,9;
3. Universidade da Beira Interior Thesis UbiThesis – 90,4%
4. Technical University of Lisbon Repository – 89,7%
5. Repositório Institucional Universidade Nova de Lisboa 89,3%

Tanto no caso brasileiro quanto no português os repositórios com maiores quantidades de documentos não estão sendo devidamente coletados pelo Google Scholar e, portanto, não tiveram taxas de indexação tão altas quanto de outros sistemas analisados. Um exemplo do lado brasileiro é o “Repositório Digital Universidade Federal do Rio Grande do Sul LUME” que se manteve na primeira posição em relação à quantidade total de documentos e à quantidade de documentos recuperados pelo Google Scholar, mas caiu para décima quinta posição em relação a taxa de indexação. Do lado português destaca-se “Repositório Aberto Universidade do Porto”, também o primeiro em quantidade de documentos e em quantidade de documentos recuperados pelo Google Scholar, mas que ficou em décimo segundo em relação a taxa de indexação, com cerca de 80% dos seus documentos indexados.

Todos os resultados levantados por este estudo podem ser observados na Tabela 1 e na Tabela 2, apresentadas a seguir.

Rankings				BRASIL	Resultados		
TV	GS	Doc	RWR	Repositório	TV	GS	Doc
1	17	32	24	Repositório Institucional Universidade Federal de Goiás	274,7	1750	637
2	9	19	18	Repositório Institucional Centro Universitário de Brasília	99,7	4860	4874
3	16	25	25	Repositório de Outras Coleções Abertas Universidade Tecnológica Federal do Paraná ROCA	99,6	2020	2027
4	33	42	45	Repositório Digital da Universidade Federal do Recôncavo da Bahia	98,1	108	110
5	21	30	43	Repositório Universidade de Caxias do Sul	97,1	712	733
6	8	18	16	Repositório Institucional Universidade Federal do Rio Grande do Norte	96,3	5340	5545
7	2	6	1	Universidade de São Paulo Biblioteca Digital de Teses e Dissertações	91,8	44000	47901
8	20	29	29	Repositório Institucional Universidade Tecnológica Federal do Paraná RIUT	90,3	772	854
9	6	10	4	Universidade Federal do Paraná Biblioteca Digital de Teses e Dissertações	88,9	14400	16181
10	22	33	30	MIRAGE Universidade Nove de Julho UNINOVE	87,6	496	566
11	3	5	6	Repositório Institucional Universidade Federal de Santa Catarina	86,6	43600	50340
12	7	11	8	Universidade de Brasília Repository	79,5	11700	14702
13	29	37	36	Repositório Digital Universidade Municipal de São	76,6	226	295

				Caetano do Sul			
14	26	35	32	Repositório Institucional Instituto Nacional de Tecnologia	76,4	325	425
15	1	1	2	Repositório Digital Universidade Federal do Rio Grande do Sul LUME	75,8	75600	99632
16	35	40	37	Repositório Institucional Centro de Tecnologia da Informação Renato Archer	73,8	93	126
17	15	24	21	Repositório Institucional Universidade Federal de Lavras	64,8	2030	3130
18	38	41	40	Repositório de Publicações Científicas da Universidade Federal do Maranhão	62,4	78	125
19	27	34	33	Repositório Universidade Estadual de Ponta Grossa	59,7	320	536
20	25	31	38	Repositório Institucional Universidade Federal de Sergipe	58,3	395	677
21	4	4	3	Alice Repository Open Access to Scientific Information Embrapa	57,9	30100	51981
22	34	39	26	Repositório Institucional Rede CEDES	57,0	105	184
23	12	16	20	Repositório Institucional Pontifícia Universidade Católica de Rio Grande do Sul	52,9	3210	6067
24	13	17	10	Repositório Institucional Fundação Oswaldo Cruz	52,7	3070	5817
25	10	14	11	Repositório Institucional Universidade Federal do Ceará	50,9	4180	8198
26	14	21	17	Repositório Institucional Universidade Federal do Rio Grande	49,5	2060	4156
27	5	3	13	Biblioteca Digital da UNICAMP Universidade de Campinas	39,9	22500	56257
28	19	22	22	Repositório Institucional Universidade Federal do Pará	39,2	1320	3364
29	37	38	31	Repositório Institucional Universidade Federal do Espírito Santo	34,6	78	225
30	28	28	23	Biblioteca Digital de Teses e Dissertações da Universidade Federal do Maranhão	33,2	308	926
31	32	36	35	Repositório Institucional Digital Instituto Brasileiro de Informação em Ciência e Tecnologia	29	116	400
32	11	12	5	Universidade Federal da Bahia Repositório Institucional	24,8	3480	13976
33	30	27	27	Biblioteca Virtual Sobre Corrupção	16,2	218	1339
34	24	15	12	Biblioteca Digital de Monografias de Graduação e Especialização Universidade de Brasília	5,1	414	8007
35	23	13	19	CBPF Index Centro Brasileiro de Pesquisas Físicas	4,6	419	9018
36	18	7	9	Acervo Digital da Universidade Estadual Paulista Júlio de Mesquita Filho	3,23	1500	46357
37	40	20	34	Biblioteca Digital de Teses e Dissertações Eletrônicas da Universidade do Estado do Rio de Janeiro	0,6	31	4686
38	39	9	7	Maxwell Pontifícia Universidade Católica do Rio de Janeiro	0,3	58	18431
39	31	2	42	Repositório Institucional UNESP Universidade Estadual Paulista Júlio de Mesquita Filho	0,2	150	69382
40	36	8	14	Biblioteca Digital da Produção Intelectual da Universidade de São Paulo	0,2	83	40297
41	41	23	39	Produção Científica da Universidade Estadual de Londrina	0,08	3	3346

42	42	26	44	Biblioteca Digital da Universidade Estadual de Maringá	0,05	1	1864
Total						282.229	603.694

Tabela 1: Resultados da análise dos repositórios brasileiros

Rankings				PORTUGAL	Resultados		
TV	GS	Doc	RWR	Repositório	TV	GS	Doc
1	34	36	36	REPAP Repositório de Administração Pública	100,7	141	140
2	33	34	29	Repositório Científico Digital da Universidade da Madeira	94,9	568	598
3	22	25	23	Universidade da Beira Interior Thesis UbiThesis	90,4	1470	1625
4	8	12	9	Technical University of Lisbon Repository	89,7	5490	6117
5	6	10	5	Repositório Institucional Universidade Nova de Lisboa	89,3	7.700	8621
6	21	24	28	Repositório Científico do Instituto Politécnico de Viseu	88,2	1630	1847
7	27	32	24	IC-online Repositório Institucional Instituto Politécnico de Leiria	88,2	743	842
8	18	21	17	Repositório Aberto Universidade Aberta	88,0	2050	2328
9	5	7	4	Repositório Universidade de Lisboa	86,4	9.190	10631
10	15	17	15	Repositório Universidade dos Açores	85,2	2470	2899
11	31	33	25	Repositório Científico do Instituto Politécnico de Santarém	84,6	649	767
12	1	1	3	Repositório Aberto Universidade do Porto	83,8	28.100	33504
13	19	20	16	Repositório Científico do Instituto Politécnico de Castelo Branco	82,8	1940	2342
14	3	4	10	Repositório Institucional Universidade Católica Portuguesa Veritati	82,4	10200	12372
15	17	19	13	Repositório Digital Universidade de Trás-os-Montes e Alto Douro	81,3	2070	2546
16	13	16	12	Biblioteca Digital Universidade Fernando Pessoa	81	2700	3331
17	4	6	7	Repositório Institucional da Universidade de Aveiro RIA	80,7	9390	11630
18	14	15	18	Repositório Científico do Instituto Politécnico de Lisboa	73,4	2550	3474
19	10	11	34	ISCTE IUL Instituto Universitário de Lisboa Repositório	72,8	4980	6832
20	2	2	1	Universidade do Minho Repositorium	72,3	20.400	28201
21	28	31	32	Repositório Hospital Prof. Doutor Fernando Fonseca	70,3	730	1037
22	20	18	22	Repositório Instituto Universitário de Ciências Psicológicas, Sociais e da Vida	67,1	1910	2846
23	12	13	14	Repositório Institucional Universidade do Algarve Sapientia	65,4	3040	4643
24	25	26	31	Repositório Centro Hospitalar de Lisboa Central	65,1	1040	1596
25	16	14	19	Repositório Científico do Instituto Politécnico do Porto	62,9	2310	3669
26	23	22	21	Repositório Científico do Instituto Nacional de Saúde	61,4	1360	2212
27	26	29	26	Repositório Institucional Escola Superior de Educação de Paula Frassinetti	58,6	771	1314
28	9	9	8	Biblioteca Digital Instituto Politécnico de Bragança	58,2	5200	8933

29	30	30	33	Repositório Científico do Centro Hospitalar do Porto	57,3	667	1163
30	24	23	20	Repositório Laboratório Nacional de Energia e Geologia	57,2	1140	1990
31	29	27	27	Repositório Institucional Hospitais da Universidade de Coimbra	43,9	669	1522
32	35	35	35	Repositório Aberto Instituto Superior Miguel Torga	42,4	137	323
33	32	28	30	Repository of the Department of Informatics University of Lisbon	39,6	568	1432
34	11	8	11	Repositório Científico da Universidade de Évora	38,8	3840	9877
35	7	3	6	Estudo Geral Repositório Digital da Universidade de Coimbra	38,1	6450	16915
36	36	37	37	Instituto Gulbenkian de Ciência Repository ARCA	14,2	18	126
37	37	5	39	Repositório Biblioteca da Faculdade de Engenharia da Universidade do Porto	0	0	12177
Total						144.281	212.422

TABELA 2: Resultados da análise dos repositórios portugueses

Discussão e considerações finais

Os resultados do estudo confirmaram em parte o problema apontado por Arlitsch e O'Brien (2012) e por Orduña-Malea e López-Cózar (2014) referente a indexação dos repositórios no Google Scholar. No entanto, os resultados dos repositórios brasileiros e portugueses foram mais positivos do que os descritos pelos autores. De forma geral, eles obtiveram taxa de indexação superior a 50%.

Contudo, na comparação dos resultados entre Brasil e Portugal verificou-se que existem diferenças em relação à taxa de indexação. Apesar de ter sido do Brasil o repositório com maior taxa de indexação, em Portugal foi relativamente maior a quantidade de repositórios com notas superiores a 50% de taxa de indexação. Uma das razões para esta diferença pode ser atribuída à natureza dos repositórios. Em Portugal, 90% dos repositórios tratavam exclusivamente da produção científica de suas instituições, enquanto que no Brasil esta proporção é de apenas 65%.

Além dos tipos de documentos, considera-se que os problemas de configuração do sistema podem afetar o desempenho da indexação dos repositórios, conforme apontou Arlitsch e O'Brien (2012). Os autores apontaram para questões relacionadas com padrão de metadados, em especial o Dublin Core. Este padrão é utilizado por no mínimo todos os sistemas que estão na plataforma DSpace, o que correspondem a 86% (68 de 79) dos repositórios analisados.

Orduña-Malea e López-Cózar (2014) ainda acrescentam que os próprios buscadores utilizados no estudo, como o Google Scholar, podem apresentar resultados contraditórios. Um dos problemas observados neste estudo referente à questão levantado por Orduña-Malea e López-Cózar, foram dois casos em que a taxa de indexação foi superior a 100%. Isto significa que existem mais documentos provenientes do repositório no Google Scholar do

que disponível no próprio repositório. Algumas hipóteses sobre este problema foram levantadas. A primeira está relacionada a exclusão de documentos do repositório, que pode se manter na base, dependendo do método utilizado. A segunda diz respeito ao período necessário para o Google Scholar atualizar sua própria base com a situação real dos repositórios. Por fim, a terceira hipótese seria a de que o buscador indexa não só os documentos depositados, mas qualquer página do repositório, conforme apontado por Ordunã–Malea e López–Cózar. Todavia, essa questão extrapola a discussão proposta por este estudo e por isso não foi explorada.

Ainda na comparação entre os dois sistemas observou-se a tendência de 5 x 50 entre a distribuição dos documentos por países. Foi identificado, nos dois países e nos dois aspectos analisados (quantidade total de documentos e quantidade de documentos recuperados pelo Google Scholar), que os cinco repositórios com as maiores quantidades reúnem cerca de 50% do total de documentos. A questão não foi aprofundada, mas o seu destaque na análise é notável.

Por fim, os resultados desse estudo demonstraram que a maioria dos repositórios brasileiros e portugueses apresenta boa taxa de indexação no Google Scholar. Além disto, alguns dos repositórios mais expressivos em quantidade de documentos, tais como o “Repositório Digital Universidade Federal do Rio Grande do Sul LUME” e o “Repositório Aberto Universidade do Porto”, apesar de já terem expressivas quantidades de documentos disponíveis podem aumentar consideravelmente a quantidade de documentos indexados pelo Google Scholar, ampliando ainda mais sua visibilidade na Internet.

Referências

AGUILLO, I. F., ORTEGA, J. L., FERNÁNDEZ, M., & UTRILLA, A. M. (2010) – Indicators for a webometric ranking of open access repositories. *Scientometrics* [Em linha]. Vol 82, Nº 3. [Consult. 1 Set. 2014]. Disponível na Internet: doi:10.1007/s11192-010-0183-y.

ARLITSCH, K.; O'BRIEN, P. (2012) – Invisible institutional repositories: addressing the low indexing ratios of IRs in Google. *Library Hi Tech*, Vol 30, Nº 1, p. 60–81.

BOAI (2002) – Budapest Open Access Initiative [Em linha]. Budapeste. [Consult. 1 Set. 2014]. Disponível na Internet: <http://www.opensocietyfoundations.org/openaccess/read>

LÓPEZ–CÓZAR, E. D, & ROBINSON–GARCÍA, N. (2012) – Repositories in google scholar metrics or what is this document type doing in a place as such? *Cybermetrics* [Em linha]. Vol 16, Nº 1. [Consult. 1 Set. 2014]. Disponível na Internet: <http://hdl.handle.net/10481/22019>.

ORDUÑA-MALEA, E., & LÓPEZ-CÓZAR, E. D. (2014) – The dark side of open access in Google and Google Scholar: the case of Latin-American repositories. *Scientometrics* [Em linha]. Vol 1, N° 18. [Consult. 1 Jul. 2014]. Disponível na Internet: doi:10.1007/s11192-014-1369-5.

RUSSEL, I. (2011) – La visibilidad de los recursos académicos. Una revisión crítica del papel de los repositorios institucionales y el acceso abierto. *Investigación Bibliotecológica* [Em linha]., Vol 25, N° 53. [Consult. 1 Jul. 2014]. Disponível na Internet: <http://www.scielo.org.mx/pdf/ib/v25n53/v25n53a7.pdf>