ECI UFMG
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

MAX CIRINO DE MATTOS

pós-graduação CIÊNCIA DA INFORMAÇÃO UFMG

DA POSSIBILIDADE DE UMA *WEB OF SCIENCE* PARA A AMÉRICA LATINA E CARIBE:

extração automática de uma base de citações do SciELO para o periódico PCI e para a Coleção Saúde Pública

Profa. Dra.	**BEATRIZ VALADARES CENDÓN**	(ECI-UFMG; orientadora)
Profa. Dra.	**MARIA CRISTINA SOARES GUIMARÃES**	(ICICT-FIOCRUZ)
Profa. Dra.	**CÍCERA HENRIQUE DA SILVA**	(ICICT-FIOCRUZ)
Prof. Dr.	**RICARDO HIROSHI CALDEIRA TAKAHASHI**	(MATEMÁTICA-UFMG)
Profa. Dra.	**MARLENE OLIVEIRA TEIXEIRA DE MELO**	(ECI-UFMG)
Profa. Dra.	**RENATA MARIA ABRANTES BARACHO PORTO**	(ECI-UFMG)
Prof. Dr.	**LUIZ CLÁUDIO GOMES MAIA**	(FUMEC) - SUPLENTE

01.10.2013

VISÃO GERAL

UTOPIA / MOTOR:

inverter a relação 99% suor + 1% análise
PARA 1% suor + 99% análise

INÍCIO [no doutorado]:

- Discussão no grupo de pesquisa: TCC Wesley R. Fernandes
  Disponibilidade no PORTAL CAPES das fontes citadas nas teses de 2005-2007 na ECI-UFMG

- Continuidade: BDTD (extração automática de PDF…)
  PIBIC – problemas de padronização das referências
  Redes Neurais: base criada manualmente - Magali R. G. Meireles

- SciELO – arquivos XML

Relevância do trabalho

- Possibilidade de…
  - estudos posteriores por área, periódico, coleção…
  - aumento da visibilidade dos periódicos do SciELO
  - automação da extração de dados estatísticos

Criação de uma metodologia que permtirá a interpretação de todos os arquivos XML do SciELO
oferecendo a possibilidade de criação de uma *Web of Science* para a América Latina, Caribe e outras Coleções do SciELO

ESTRUTURA DA APRESENTAÇÃO

• Introdução	(Cap. 1)	1’ – 11’
• Fundamentação teórica	(Cap. 2)	11’ – 25’
• Metodologia	(Cap. 3)	25’ – 37’
• FASE I – Dados Cadastrais		
• Resultados	(Cap. 4)	
• Análise da FASE I	(Cap. 7)	
• FASE II – Base de Citações	(Cap. 3)	
• Análise das FASES I e II	(Cap. 7)	
• Resultados para a PCI	(Cap. 5)	37’ – 42’
• Resultados para a CSP	(Cap. 6)	42’ – 46’
• Análise da FASE II	(Cap. 7)	46’ – 48’
• Conclusões e recomendações	(Cap. 8)	48’ – 50’

INTRODUÇÃO

1. INTRODUÇÃO 14
1.1 Justificativa 15
1.2 O protótipo 17
1.3 Problema e questão de pesquisa 23
1.4 Objetivos 23
1.5 Premissas e limites da pesquisa 23
1.6 Estrutura do trabalho 24

MAPAS FRAGMENTADOS

- Garfield (processamento manual)
- PORTAL CAPES
- ISI seletivo (fragmentação local)

ARMAZENAMENTO MAGNÉTICO

- Grandes volumes de dados
- Automação
- Nascimento do ISI - WoS

ÍNDICES LOCAIS

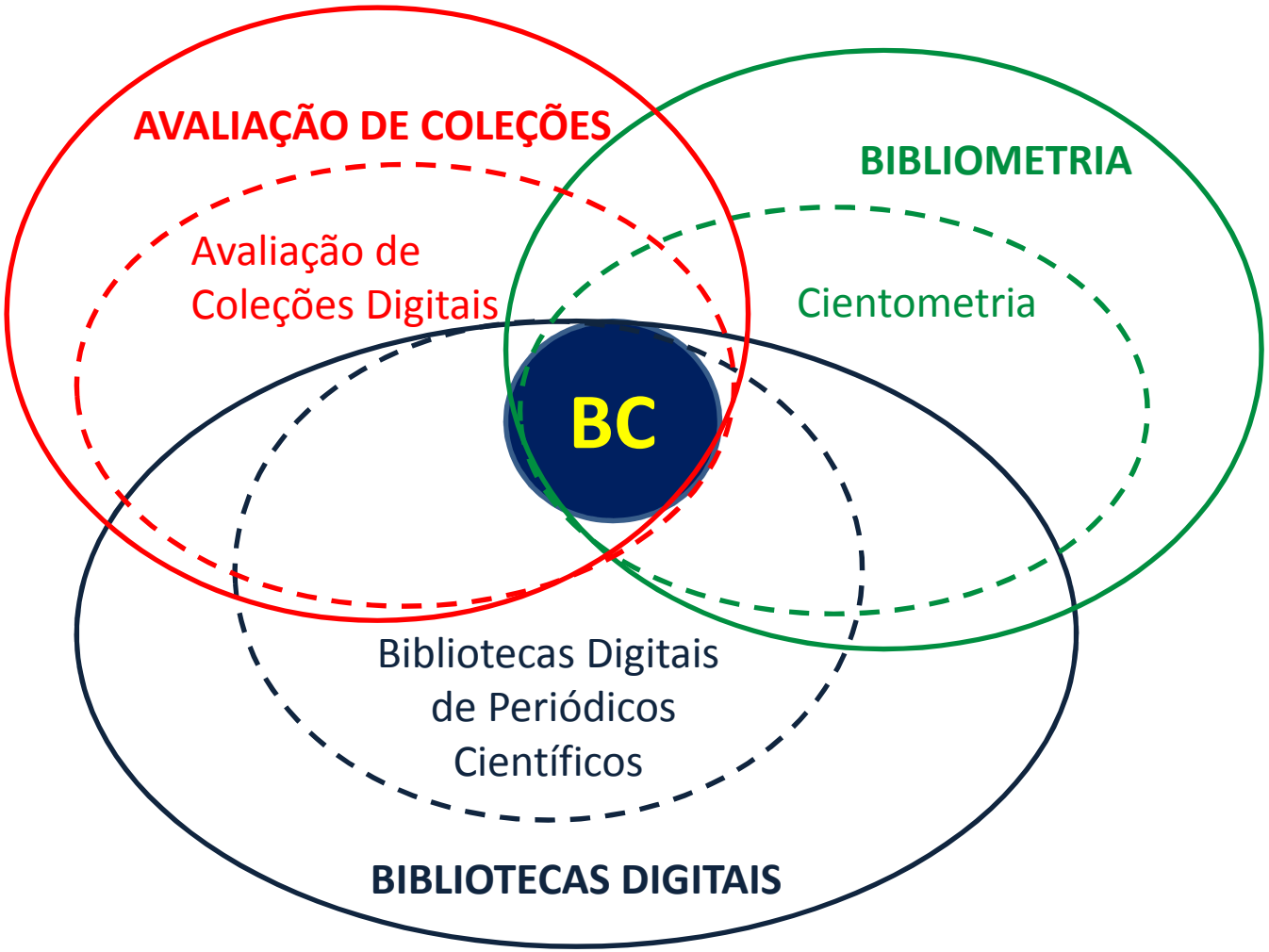
- Importância
- Políticas, investimentos, avaliação

SciELO

Extração automática (XML)

INTRODUÇÃO

FIGURA 1 – Fundamentos para estudo das bases de citações



Fonte: Desenvolvida pelo autor

Estruturação da fundamentação teórica

(CAP. 2)

MAPAS FRAGMENTADOS

- Garfield (processamento manual)
- PORTAL CAPES
- ISI seletivo (fragmentação local)

ARMAZENAMENTO MAGNÉTICO

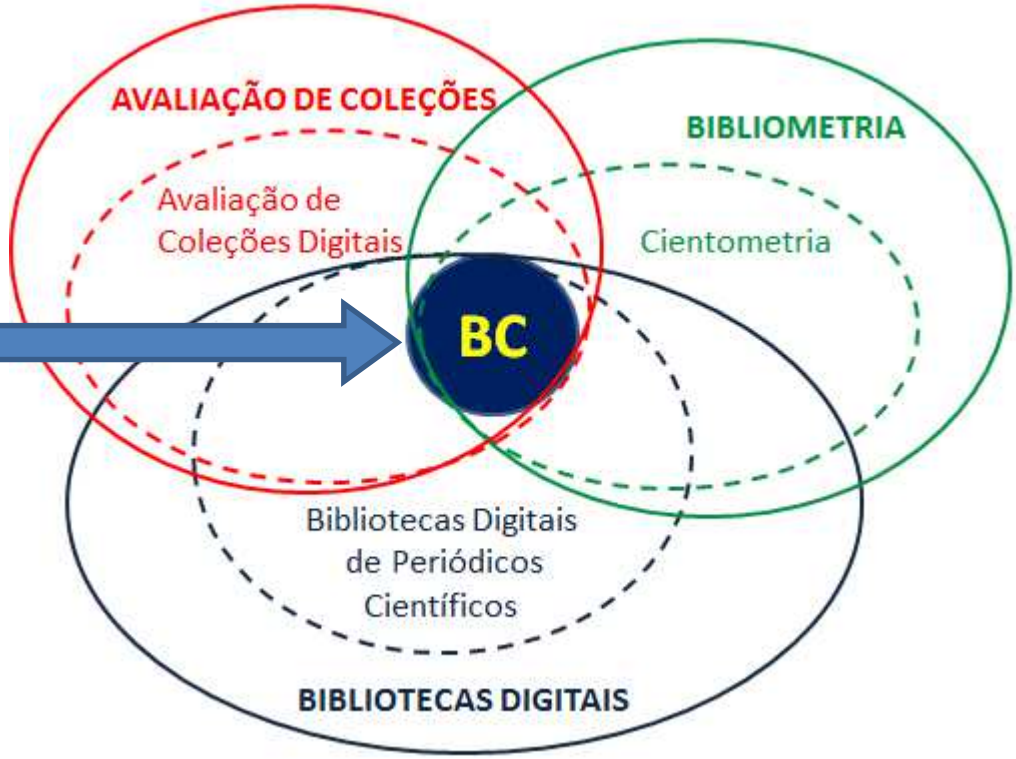
- Grandes volumes de dados
- Automação
- Nascimento do ISI - WoS

ÍNDICES LOCAIS

- Importância
- Políticas, investimentos, avaliação

SciELO

Extração automática (XML)



INTRODUÇÃO

Justificativa

- Medir publicações científicas não cadastradas no ISI é importante, porém há carência de bases de dados nesses moldes
- SciELO: iniciativa para contemplar essas publicações (MENEGHINI, 1998)

- Áreas de conhecimento em que a tradição é de disseminação local (CENDÓN, GUIMARÃES, SILVA, OLIVEIRA, MATTOS, SANTANA e FERNANDES, 2012)

- Processos de gestão de atividades científicas deformados pela falta de índices locais (GUIMARÃES, SILVA, SANTANA, BRAGA, BOCHNER e GOLDBAUM, 2011)

INTRODUÇÃO

Justificativa

- Dificuldade: processar um enorme volume de dados
- Esforços datam de 1927 (mapeamento fracionado) (GARFIELD, 1972)

- Usar a tecnologia como recurso essencial de desenvolvimento da CI

Protótipo

- Integração CC e CI:
  Lourenço (2005) – MER e MTD-BR
  Bohmerwald (2005) – usabilidade e estudos de usuários
  Maia (2013) – MER e Categorias de Ranganathan

INTRODUÇÃO

Pergunta

É POSSÍVEL A AUTOMAÇÃO DO PROCESSO DE CRIAÇÃO DE UMA BASE DE CITAÇÕES PARA OS PERIÓDICOS DO SciELO?

Objetivos

- Descrição de uma metodologia para a criação de uma base de citações atualizada automática e continuamente a partir dos periódicos cadastrados no SciELO

  - Desenvolver um protótipo para testar a metodologia
  - Validar o protótipo para testar sua aplicação prática
  - Identificar fatores limitadores e propor alternativas

INTRODUÇÃO

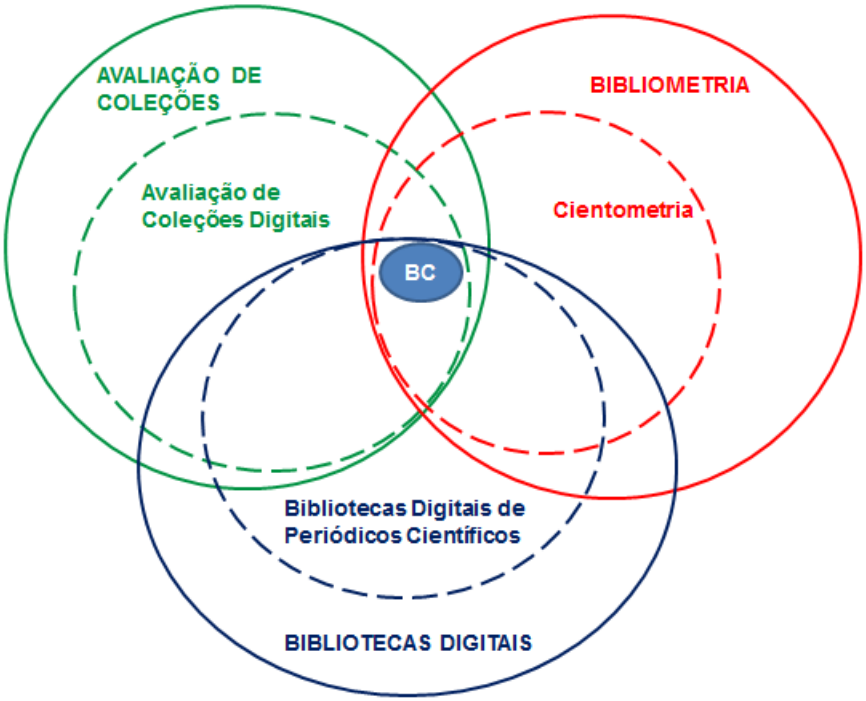
Premissas e limites

- Não se pretendeu corrigir o conteúdo dos arquivos XML
- Não foram usados critérios para filtragem de informações
- Não integra o escopo do trabalho a análise específica de nenhuma área
- O processo depende dos padrões identificados para acesso aos dados do SciELO

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

11’

2. PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA 25
2.1 Periódicos científicos e a internet: bibliotecas digitais 25
2.2 Bibliometria e Cientometria: criação e uso da *Web of Science* 29
2.2.1 *Web of Science* 31
2.2.2 Aplicações práticas 35
2.3 Bases de citações locais e o uso do SCIELO 37
2.3.1 SCIELO .....
2.4 Outras aplicações: a avaliação de coleções d
2.4.1 Avaliação de coleções: o acervo do POF



PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Periódicos científicos e a Internet: bibliotecas digitais

- "A comunicação [científica] situa-se no próprio coração da ciência" (MEADOWS, 1999)

- Internet: visibilidade a todos os periódicos
- Preços: incentivo à disseminação de periódicos de acesso aberto
- SciELO (meados 1990)
  (MENEGHINI E SILVA, 2012, informação verbal)

- Bibliotecas digitais:
  - Expansão do número de usuários e abrangência
  - Gerenciamento por LOG (CUNHA, 2009; DIAS, 2012)
  - Falta de literatura específica (MARCONDES ET AL., 2006)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bibliometria e Cientometria: criação e uso da *Web of Science*

- Price (Lotkka, Bradford, Zipf): leis cientométricas; "mapas de ciência"
- "Mapas": evidenciam importância e cobertura das revistas
- "Mapas": embasamento do projeto do ISI
  (SANTOS E KOBASHI, 2009)

- VINITI – *All-Union Institut for Science and Technical Information*
- Academia de Ciências da ex-URSS
- 1969: Nalimov e Mulchenko – termo "cientometria" no título
- Nalimov (cibernética), Styazhkin e Vledutsiv em 1959:
  "As informações científicas e técnicas como uma das tarefas da cibernética"
  (VANTI, 2011)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bibliometria e Cientometria: criação e uso da *Web of Science*

- Garfield percebeu a possibilidade de uso das referências citadas como elementos de recuperação assim como palavras-chave
- Base para a criação da WoS
  (VANTI, 2011)

- WEB OF SCIENCE
  - +12.000 periódicos de grande impacto no mundo inteiro
  - +150.000 conferências
  - + 250 disciplinas
  - A partir de 1.900
    (site WoS; 12 ago. 2013)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bibliometria e Cientometria: criação e uso da *Web of Science*

- ISI: "rei" absoluto
- "Candidatos ao trono": limitação uma/poucas áreas; não empreendem esforços para prover índices de citação nos moldes do ISI (ADAM, 2002)

	Periódicos	Fascículos		Artigos		Citações Concedidas		Citações Recebidas	
8 áreas	904	29.728	100,00%	373.370	100,00%	9.823.290	100,00%	729.546	100,00%
CIENCIAS AGRARIAS	96	3.133	10,54%	58.057	15,55%	1.383.257	14,08%	162.645	22,29%
CIENCIAS BIOLOGICAS	94	3.472	11,68%	49.493	13,26%	1.382.040	14,07%	102.499	14,05%
CIENCIAS DA SAUDE	290	12.565	42,27%	151.513	40,58%	3.769.443	38,37%	333.465	45,71%
CIENCIAS EXATAS E DA TERRA	67	2.052	6,90%	28.331	7,59%	818.982	8,34%	31.946	4,38%
CIENCIAS HUMANAS	226	4.232	14,24%	44.425	11,90%	1.332.891	13,57%	61.560	8,44%
CIENCIAS SOCIAIS APLICADAS	148	2.405	8,09%	20.573	5,51%	633.131	6,45%	23.364	3,20%
ENGENHARIAS	70	1.545	5,20%	18.208	4,88%	427.500	4,35%	12.677	1,74%
LINGUISTICA LETRAS E ARTES	23	324	1,09%	2.770	0,74%	76.046	0,77%	1.390	0,19%

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bibliometria e Cientometria: criação e uso da *Web of Science*

(Interesse: criação do ISI -> criação da BC do SciELO)
GARFIELD (1972; 1979; 1992; 1995)

- Mapas fragmentados (esforço manual)
- Uso de meio magnético
- Amostra: out/dez 1969 (2.200 periódicos; 1.000.000 citações)
- Listagens:
  - Frequência de citações
  - Estatística dos periódicos citados
  - Estatística dos periódicos citantes

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

FIGURA 6 – Frequências de citações

ITEM NO	CITED JOURNAL	TOTAL	NUMBER OF TIMES CITED									
			1969	1968	1967	1966	1965	1964	1963	1962	1961	1960	REST
00243	ACTA PATH JAP	36	1	3	3	4	6	7	3	·	·	3	6
00244	ACTA PATH MICROBIOL	736	29	69	87	59	56	59	44	48	20	31	234
00909	AM J ANAT	637	7	27	37	56	32	41	15	26	15	21	360
00910	AM J BOT	1171	13	74	87	68	73	66	57	57	47	49	580
00911	AM J CANCER	103	·	·	·	·	·	·	·	·	·	·	103
00912	AM J CARDIOL	1238	73	201	199	247	134	70	78	66	53	73	44
03591	CAN J SOIL SCI	33	·	2	1	4	1	2	6	6	3	2	6
03592	CAN J SURG	61	2	6	4	3	13	11	3	3	5	1	10
03593	CAN J TECH	3	·	·	·	·	·	·	·	·	·	·	3
03594	CAN J ZOOL	356	46	38	40	28	24	20	19	29	17	22	73
08990	ISRAEL J AGR RES	29	1	1	7	·	1	8	7	2	1	·	1
08991	ISRAEL J BOT	16	·	·	3	5	4	3	1	·	·	·	·
08992	ISRAEL J CHEM	91	14	25	18	10	11	6	7	·	·	·	·
09651	J INVEST DERM	695	24	78	81	69	65	46	30	31	34	22	215
09652	J IOWA MED SOC	13	·	·	·	5	·	·	1	2	·	·	5
09653	J IRISH MED ASS	16	1	3	4	3	3	·	·	·	·	·	2
13390	P CALIF ACAD SCI	18	·	·	1	4	3	·	1	·	·	·	9
13391	P CAMBRIDGE PHIL SOC	389	8	22	23	11	12	9	13	11	17	3	260
19755	Z ANGEW CHEM	47	·	·	1	·	·	1	·	·	1	1	43
19756	Z ANGEW ENT	35	·	·	1	1	4	2	4	1	1	5	16
19757	Z ANGEW GEOL	49	2	7	5	8	5	5	4	4	2	1	6
19758	Z ANGEW MATH	10	1	·	1	·	1	·	1	1	1	·	4

Fig. 1. Journal citation frequencies. The data show the total number of times each journal was cited during the last quarter of 1969 and the distribution by publicatio date of the particular issues cited. The journals shown were taken from a list of m than 20,000 items (journals, books, reports, theses, and so on) cited during the quarter of 1969 in journals covered by the SCI.

Fonte: Garfield, 1972, p.528

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

CITED JOURNAL / CITING JRNL	TOTAL	NUMBER OF TIMES CITED 1969	1968	1967	1966	1965	1964	1963	1962	1961	1960	REST
J LINN SOC BOT	17*	0	3	0	3	1	0	0	0	1	1	8
NEW PHYTOL	5	0	1	0	1	0	0	0	0	1	0	2
ALL OTHER (10)	12	0	2	0	2	1	0	0	0	0	1	6
J LIPID RES	902*	33	108	109	98	133	121	75	83	60	58	24
BIOC BIOP A	83	5	8	16	6	13	15	4	6	5	1	4
BIOCHEM J	39	0	4	7	3	6	7	3	3	4	1	1
J BIOL CHEM	39	0	6	7	3	8	3	6	2	2	2	0
J CHROMAT	30	2	3	4	4	2	6	0	4	2	1	2
J CLIN INV	29	1	3	3	3	7	2	2	3	3	1	1
J LIPID RES	28	5	5	4	2	4	1	2	0	3	1	1
P SOC EXP M	25	0	3	1	3	4	3	2	5	1	1	2
MILIT MED	5	0	0	2	0	0	0	1	1	0	1	0
NY ST J MED	5	0	1	1	2	0	0	0	1	0	0	0
ALL OTHER (123)	219	6	19	16	29	43	27	23	15	15	23	3
J LOND MATH SOC	173*	16	19	19	11	8	10	2	8	4	8	68
J LOND MATH	71	8	12	7	5	4	4	0	4	2	4	21
T AM MATH S	11	2	1	2	1	0	0	0	1	0	0	4
P CAMB PHIL	8	1	0	3	0	0	0	0	0	0	0	4

Fig. 2. Statistics on cited journals. The data on cited journals show the total number of times each journal was cited in the last quarter of 1969 and the distribution by publication date of cited issues. For each cited journal, the figure identifies all other journals ("citing journals") that referred to it five times or more during the quarter year (and the distribution of these citations by publication date of cited issues). Journals that referred to the cited journal less than five times are grouped as "all other." The data were taken from a complete list of journals cited during the last quarter of 1969 in journals covered by the SCI.

FIGURA 7 – Estatísticas dos periódicos citados

Fonte: Garfield, 1972, p.529

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

SOURCE JOURNAL / REFERENCE JOURNAL	TOTAL	NUMBER OF TIMES CITED 1969	1968	1967	1966	1965	1964	1963	1962	1961	1960	REST
J LIBR AUT ==========>	168*	15	50	35	19	8	14	5	6	3	1	12
PROGRAM	15	3	6	5	1	0	0	0	0	0	0	0
J LIBRARY AUTOMATION	7	3	4	0	0	0	0	0	0	0	0	0
ALL OTHER (120)	146	9	40	30	18	8	14	5	6	3	1	12
J LIPID RES ==========>	313*	15	36	35	26	25	29	20	10	15	15	87
J BIOL CHEM	43	2	6	3	6	1	3	1	0	1	1	19
J LIPID RES	28	5	5	4	2	4	1	2	0	3	1	1
BIOCHIM BIOPHYS ACTA	19	1	3	1	3	2	3	1	0	2	2	1
BIOCHEM J	13	1	1	3	0	0	0	1	2	1	1	3
J AMER CHEM SOC	12	0	0	0	0	0	0	0	0	1	0	11
BIOCHEMISTRY	9	0	3	0	2	1	2	1	0	0	0	0
J CHEM SOC LONDON	5	0	0	1	0	0	0	0	0	1	0	3
J CHROMATOGR	5	0	1	1	1	0	1	0	1	0	0	0
METHODS ENZYMOL	5	0	0	0	0	0	0	0	1	0	0	4
ALL OTHER (89)	124	5	12	19	10	8	16	10	2	5	6	31
J LOND MATH ==========>	743*	21	44	58	52	42	34	31	38	29	33	361
J LOND MATH	52	6	12	5	5	4	3	0	2	1	3	11
P LONDON MATH SOC	37	4	5	7	1	2	2	0	1	1	0	14

Fig. 3. Statistics on citing journals. The data on each citing journal ("source journal") show the total number of references each journal contained in the last quarter of 1969 and the distribution of that total by publication date of journal issues referred to. For each citing journal, the figure identifies all journals ("reference journal") cited five times or more during the quarter year (and gives the distribution by publication data of cited issue). Journals cited less than five times are grouped as "all other." The data are taken from a list of journals processed for the SCI during the last quarter of 1969.

FIGURA 8 – Estatísticas dos periódicos citantes

Fonte: Garfield, 1972, p.530

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bibliometria e Cientometria: criação e uso da *Web of Science*

(Interesse: criação do ISI -> criação da BC do SciELO)
Base de citação (GARFIELD):

- Potencial de gerenciamento de coleções de periódicos
- Principal aplicação: avaliação de pesquisas e políticas científicas
- Definição de políticas de investimento e avaliação de desempenho
- Alerta para a cobertura seletiva do ISI: sugeriu a criação de um SCI para a América Latina

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bibliometria e Cientometria: criação e uso da *Web of Science*

- WoS x *Google Scholar* (WINTER, ZADPOOR E DODOU, 2013)
  - Acesso restrito X público
  - Indexação seletiva X coleta automática (frágil)
    Jacsó (2005; 2008), Cathcart e Roberts (2005);
    Donlan e Cooke (2005); Vine (2006) e Wleklinski (2005)

- Evolução do *Google Scholar*
  - Química, Física, Economia, Medicina (HARZING, 2013)
  - Perfis de +30.000 pesquisadores (RADICCHI E CASTELLANO, 2013)
    - Críticas aos resultados (PRATHAP, 2013)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bibliometria e Cientometria: criação e uso da *Web of Science*

- Colaboração científica
  - Coautoria (SCHUBERT E BRAUN, 1990)
  - Cooperação Turquia/Malásia: combustíveis energéticos (KUMAR E JAN, 2013)

- Estudos de gênero
  - Produção feminina iraniana (NOURMOHAMMADI E HODAEI, 2013)
  - Preferência no uso de gráficos e tabelas entre homens e mulheres (HARTLEY E CABANAC, 2013)
  - Nanotecnologia – JCR 2005 e 2007 (SOTUDEH E KHOSHIAN, 2013)
    - Uso do EXCEL e SPSS (99% suor...)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bases de citações locais e o uso do SciELO

- Negishi, Sun e Shigi (2004)
  - Criação de um índice de citações para periódicos japoneses *Citation Database for Japanese Papers* (CJP)
  - Importância da criação de índices de língua não inglesa
    - *Chinese Science Citation Database*
    - *China Scientific and Technical Papers and Citations Database*
    - *Chinese Social Science Citation Index*

- Zibareva e Parmon (2012) : *Russian Science Citation Index*

- Suleimenov *et al*. (2009) : *Kazakh Science Citation Index* (dados de aproximadamente 30.000 pesquisadores)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bases de citações locais e o uso do SciELO

- América Latina e Caribe
  - Krauskopf *et al*. (1995) : ISI entre 1981 e 1993
    Argentina, Brasil, chile, Colômbia, Costa Rica, Cuba, Jamaica, México, Peru e Venezuela (recebido em 10 mai. 1995)
  - Collazo-Reyes (2013): português passou a ser a segunda língua – atrás apenas do inglês – na produção científica registrada no ISI para países da América Latina e Caribe
  - Análise bibliométrica da produção científica sobre AIDS na América Latina e Caribe
    (MACIAS-CHAPULA, RODEA-CASTRO E NARVAEZ-BERTHELENOT, 1998)
    - 82 artigos no **SCI (3,93%)**, 272 no PERIODICA (13,05%), 765 no LILACS (36,69%) e 966 no AIDSLINE (46,33%)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bases de citações locais e o uso do SciELO

- O uso de indicadores provenientes do SciELO contribui para a adequação de critérios de avaliação da produção científica nacional (MUGNANI, 2006)

- "a metodologia do Projeto [SciELO] é diferenciada do conjunto de informações levadas à Internet, por obedecer a **rígido controle** de avaliação, seleção e operacionalidade"
(VASCONCELLOS, 1999)

- Guedes (2012) – narração da história do SciELO (entrevistas)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Bases de citações locais e o uso do SciELO

- Obtenção dos dados estatísticos de forma manual
  - Goldenberg et al. (2007) – Acta Cirúrgica Brasileira
  - Solano e Valdivea (2003) – AIDS 1997 a 2003
  - Población e Goldenberg (2001) – Acta Cirúrgica Brasileira
- **A importância do SciELO é inegável**

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Outras aplicações: avaliação de coleções digitais

- O que o acervo possui e não deveria possuir;
  e o que não possui mas deveria possuir…
  (LANCASTER, 1996)

- Acervo do PORTAL CAPES
  - 45% dos artigos citados estavam disponíveis no PORTAL CAPES
  - Análise bibliométrica como critério para gestão do acervo (FERNANDES E CENDÓN, 2010)

  - Histórico de criação do PORTAL CAPES e composição do acervo (CORREA ET AL., 2008; ALMEIDA, GUIMARÃES E ALVES, 2010; FERNANDES, 2012)

PERIÓDICOS CIENTÍFICOS, BIBLIOTECAS DIGITAIS E CIENTOMETRIA

Outras aplicações: avaliação de coleções digitais

- Estudos sobre a coleção do PORTAL CAPES
  - Avaliação a partir da percepção do usuário (MAIA, 2005; CUNHA, 2009; SANTANA e PEIXOTO, 2010; FERNANDES, 2012; MAIA e CENDÓN, 2012)
  - Avaliação a partir da existência das referências citadas (OLIVEIRA e ODDONE, 2007; FERNANDES, 2009; FERNANDES e CENDÓN, 2010; SANTANA e PEIXOTO, 2010)
- Limitações:
  quantitativa e geográfica (universo de instituições)
  quantitativa e temática (áreas de conhecimento)

MANUAL

METODOLOGIA

25’

3. METODOLOGIA 48
3.1 Definição das ferramentas para desenvolvimento 50
3.2 FASE I: Obtenção dos dados do SCIELO 50
3.2.1 Identificação dos periódicos de uma Coleção do SCIELO 51
3.2.2 Obtenção dos dados estatísticos anuais de cada periódico 54
3.2.3 Vinculação da(s) área(s) de conhecimento de cada periódico 55
3.2.4 Criação do Módulo “Dados Cadastrais” do banco de dados 56
4 RESULTADOS DA FASE I: COLEÇÕES DO SCIELO EM NÚMEROS 65
4.1 Resumo dos dados estatísticos: o Módulo Dados Cadastrais em números 65
4.2 Identificação dos periódicos e obtenção dos dados estatísticos 66
4.3 Vinculação às áreas de conhecimento 73
7. ANÁLISE DOS RESULTADOS 109
7.1 Análise da obtenção dos dados do SCIELO 109
7.2 Análise da criação da base de citações 116
3.3 FASE II: Criação da Base de Citações 58
3.3.1 Levantamento dos arquivos XML de cada periódico 58
3.3.2 Interpretação dos arquivos XML do SCIELO 59
3.3.3 Criação do Módulo “Base de Citações” do banco de dados 63
3.4 A seleção da amostra 64

METODOLOGIA

Caracterização da pesquisa

- Pesquisa aplicada ou exploratória (LAKATOS E MARCONI, 2007)

  Demonstrar a viabilidade de um determinado programa ou técnica como uma solução em potencial para problemas práticos

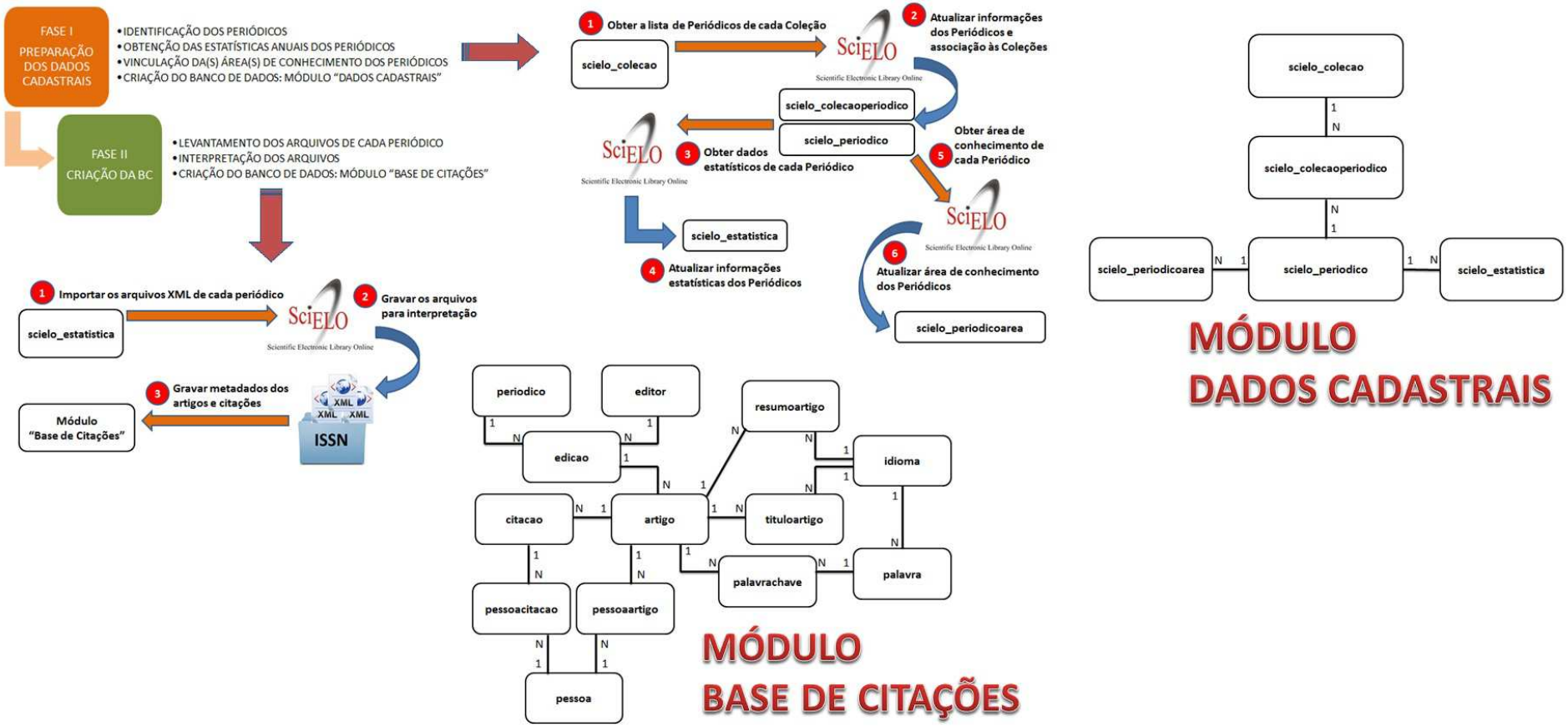
- Híbrida (CRESWELL E CLARCK, 2011)

  Utiliza métodos qualitativos (abstração, generalização - metodologia) e quantitativos – base de citações associada à bibliometria

- Técnica de modelagem relacional (CODD, 1969; 1970)
- MER (CHEN, 1976; 2002)

METODOLOGIA

FIGURA 9 – Visão geral da metodologia: passos e modelagem de dados simplificada

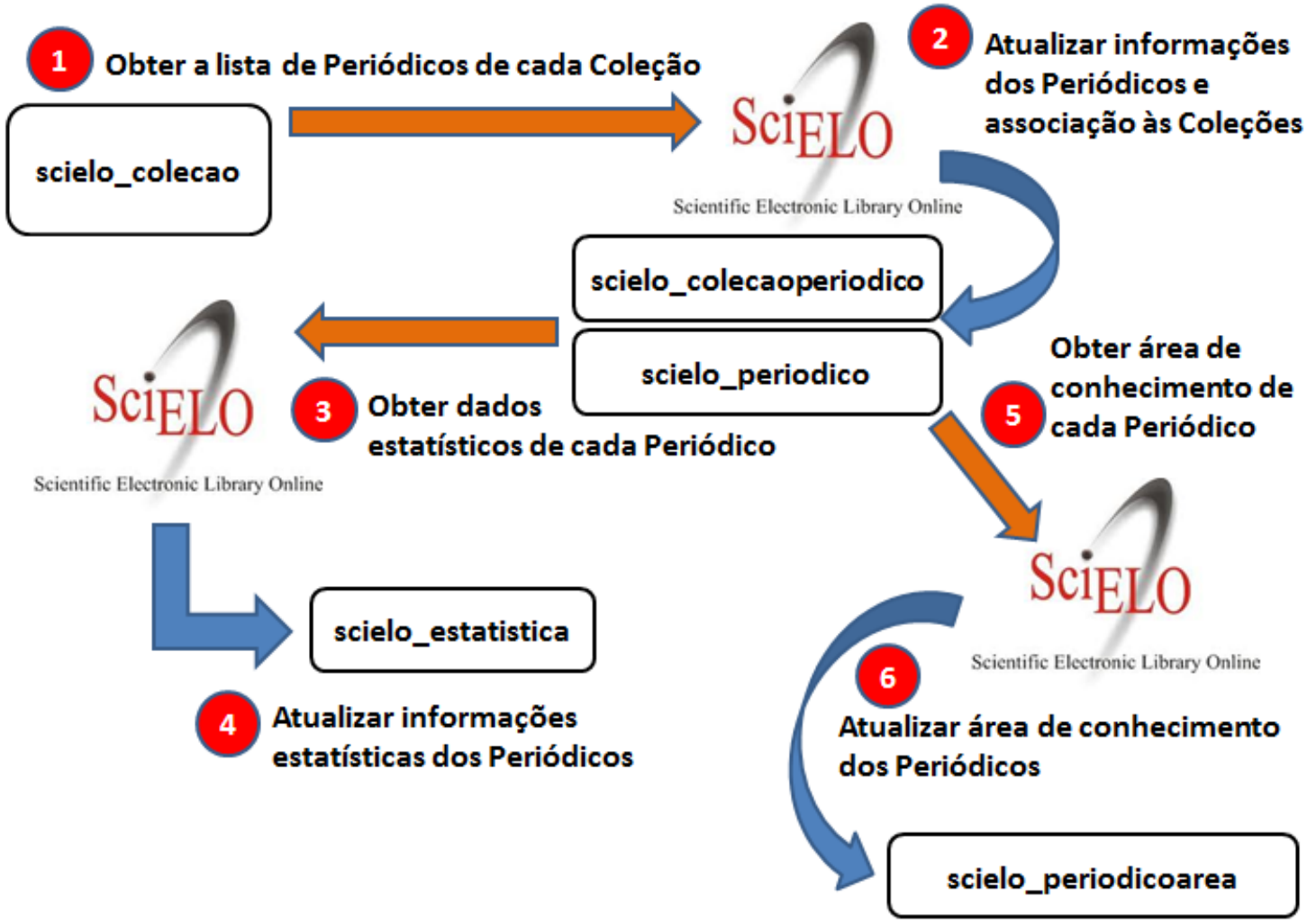


Fonte: Desenvolvida pelo autor

METODOLOGIA

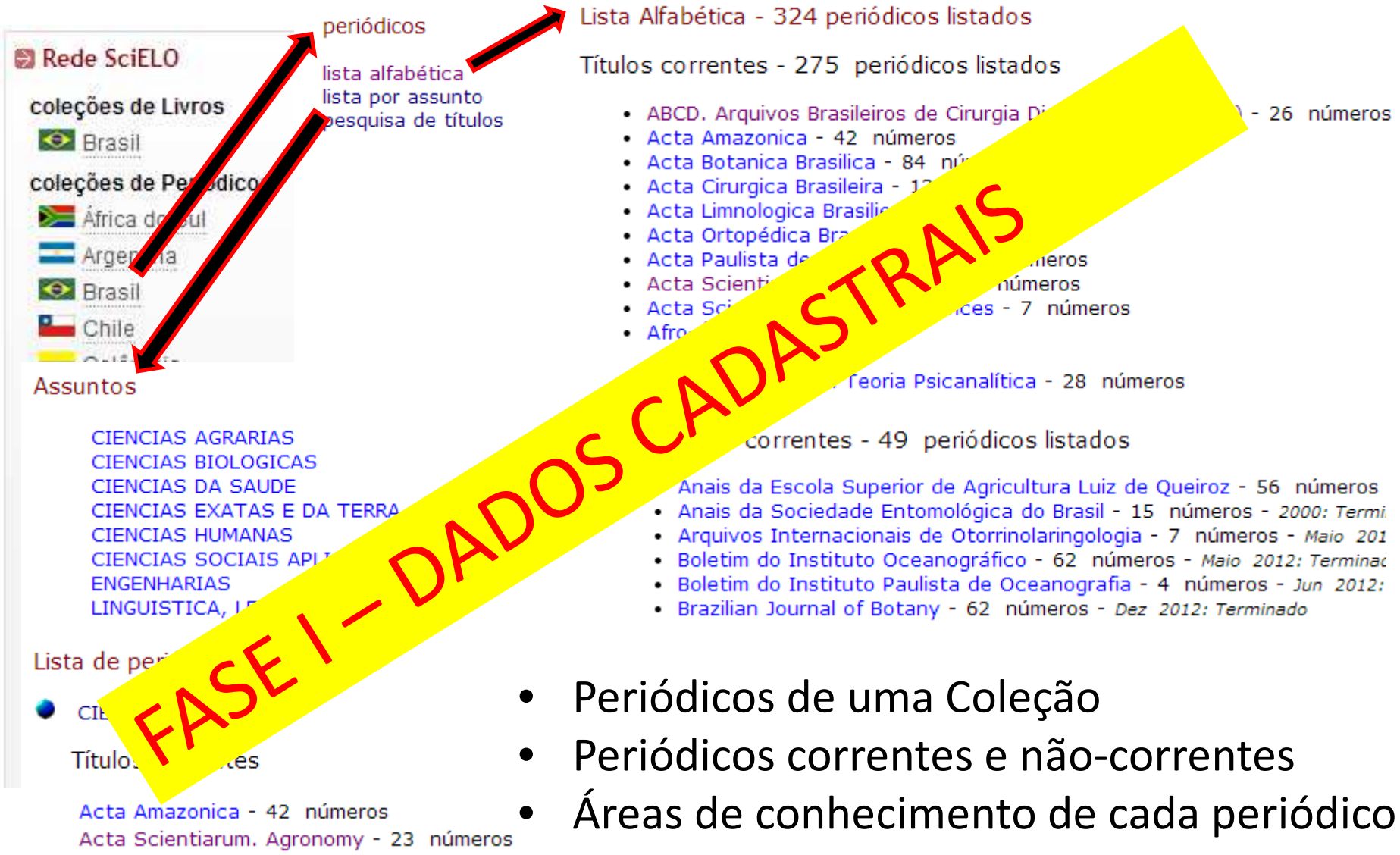
FASE I – Dados Cadastrais

FIGURA 3 – Processo automático de preparação dos dados cadastrais dos periódicos do SciELO



Fonte: desenvolvida pelo autor

METODOLOGIA



- Periódicos de uma Coleção
- Periódicos correntes e não-correntes
- Áreas de conhecimento de cada periódico

METODOLOGIA

- Dados fonte de cada periódico

Lista Alfabética - 324 periódicos listados

Títulos correntes - 275 periódicos listados

- ABCD. Arquivos Brasileiros de Cirurgia Digestiva (São Paulo) - 26 números
- Acta Amazonica - 42 números
- Acta Botanica Brasilica - 84 números
- Acta Cirurgica Brasileira - 126 números
- Acta Limnologica Brasiliensia - 11 números
- Acta Ortopédica Brasileira - 67 números
- Acta Paulista de Enfermagem - 48 números
- Acta Scientiarum. Agronomy - 23 números
- Acta Scientiarum. Animal Sciences - 7 números
- Afro-Ási 3 números
- Ágor Estudos em Teoria Psicanalítica - 28 números



iarum.
ny
1807-8621

FASE I – DADOS CADASTRAIS

Atualizado em
Agosto 20, 2013

english
español

- sobre nós
- corpo editori
- instruções aos
- assinaturas
- estatísticas

tações de rev
- os fonte
- Fator de impacto em um perí
- Fator de impacto em um perí
- Vida média
- Citações recebidas
- Citações concedidas

Lista de dados fonte

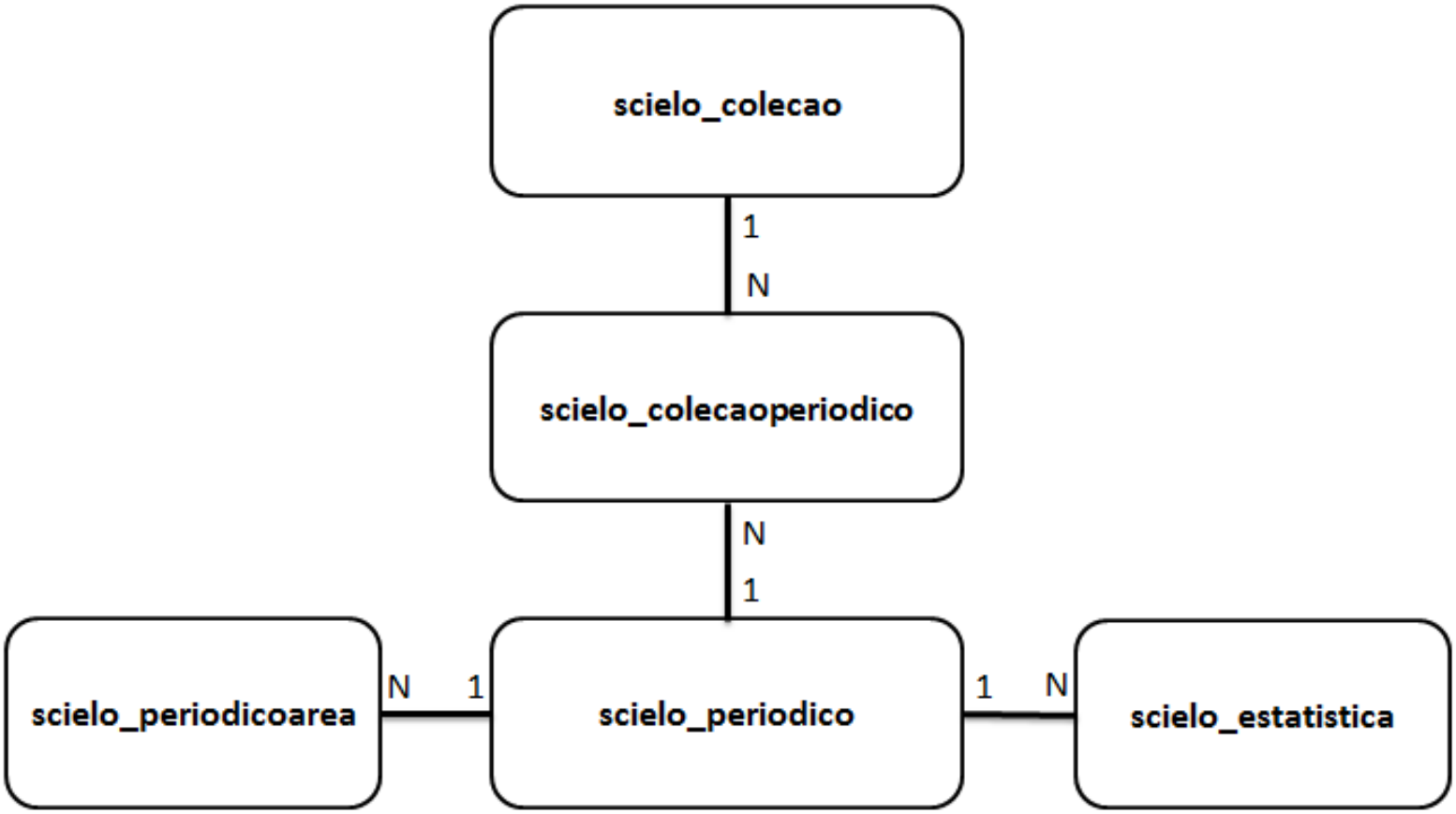
ara selecionar a coluna de ordenação
a a ordem corrente

titulo da revista/ano	n. de fascículos	n. de artigos	n. de citações concedidas	n. de citações recebidas	média de artigos por fascículo
Acta Sci., Agron	**20**	**430**	**10599**	**1438**	**21.50**
2013	2	30	1002	165	15.00
2012	4	60	1476	276	15.00
2011	4	100	2452	372	25.00
2010	4	100	2311	302	25.00
2009	4	100	2307	239	25.00
2008	2	40	1051	84	20.00
total	**20**	**430**	**10599**	**1438**	**21.50**

METODOLOGIA

FASE I – Dados Cadastrais

FIGURA 12 – Módulo “Dados Cadastrais”

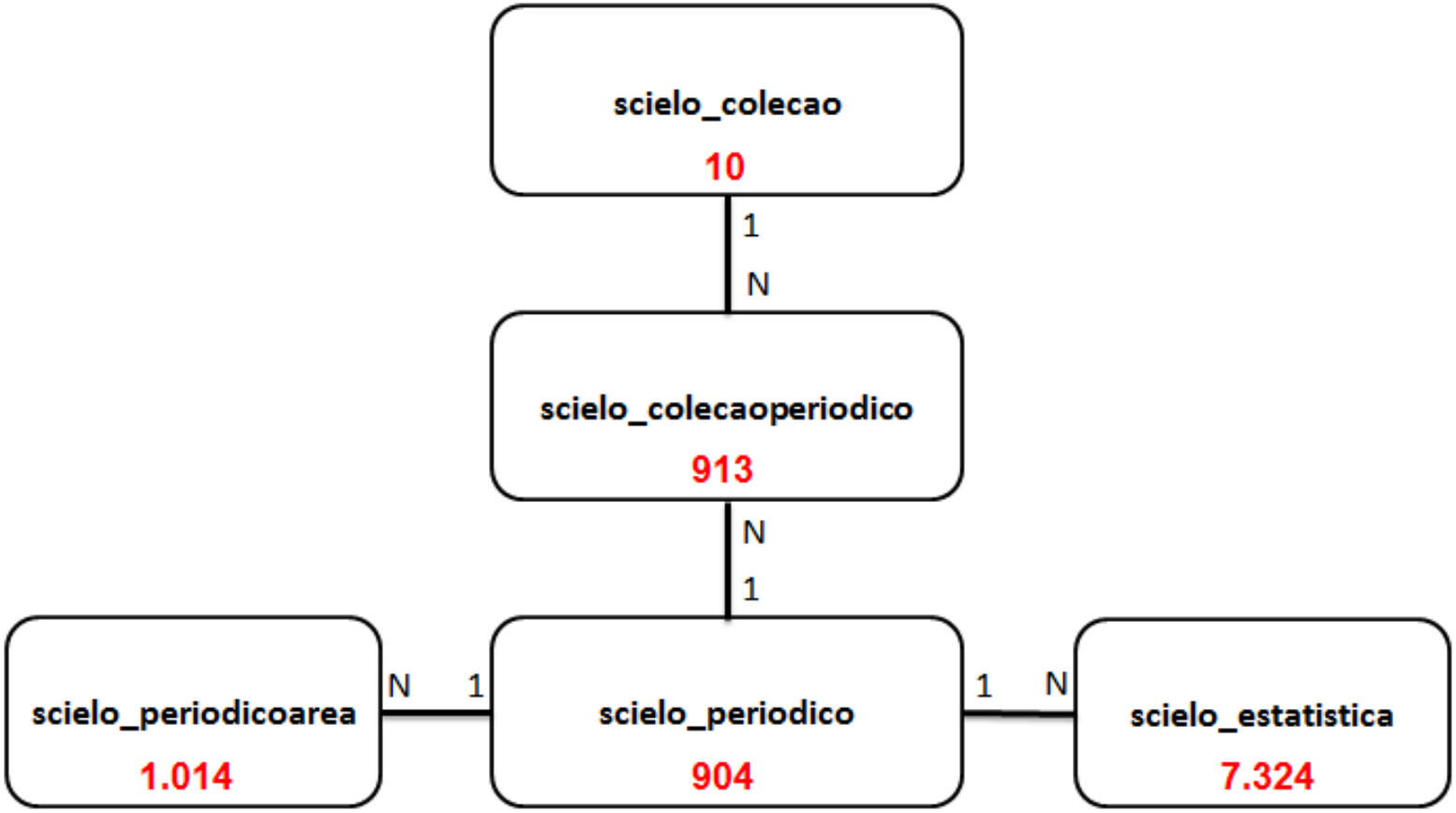


Fonte: desenvolvida pelo autor

METODOLOGIA

FASE I – Dados Cadastrais – Resultados (CAP. 4)

FIGURA 19 – Módulo “Dados Cadastrais” em números



Fonte: desenvolvida pelo autor

METODOLOGIA

FASE I – Dados Cadastrais – Resultados (CAP. 4)

- 904 periódicos
- 895 vinculados a somente 1 Coleção
- 9 encontrados em 2 Coleções
  - 1 delas sempre a CSP

MESMO ISSN EM DUAS COLEÇÕES
DADOS CADASTRAIS DIFERENTES?

ii

METODOLOGIA

FASE I – Dados Cadastrais – Resultados (CAP. 4)

FIGURA 57 – Lista de dados fonte na Coleção Saúde Pública: ISSN 0124-0064

Revista de Salud Pública
Journal of Public Health

Revista de Salud Pública
ISSN 0124-0064

Data do último processamento 11-06-2013

Lista de dados fonte

∇ - clique para selecionar a coluna de ordenação
Δ - indica a ordem corrente

titulo da revista/ano Δ	n. de fascículos ∇	n. de artigos ∇	n. de citações concedidas ∇	n. de citações recebidas ∇	média de artigos por fascículo ∇
Rev. salud pública	**47**	**575**	**13949**	**302**	**12.23**
2012	4	50	1228	31	12.50
2011	6	87	2099	47	14.50
2010	8	100	2625	67	12.50
2009	6	89	2208	42	14.83
2008	6	88	2190	47	14.67
2007	4	54	1172	19	13.50
2006	5	52	1003	21	10.40
2005	3	28	646	13	9.33
2004	4	22	660	11	5.50
2003	1	5	118	4	5.00
total	**47**	**575**	**13949**	**302**	**12.23**

Fonte: SCIELO, 2013[59]

ACESSO EM: 13 JUN. 2013

FIGURA 58 – Lista de dados fonte na Coleção Colômbia: ISSN 0124-0064

Revista de Salud Pública
Journal of Public Health

Revista de Salud Pública
ISSN 0124-0064

Data do último processamento 11-06-2013

Lista de dados fonte

∇ - clique para selecionar a coluna de ordenação
Δ - indica a ordem corrente

titulo da revista/ano Δ	n. de fascículos ∇	n. de artigos ∇	n. de citações concedidas ∇	n. de citações recebidas ∇	média de artigos por fascículo ∇
Rev. salud pública	**57**	**644**	**15759**	**1135**	**11.30**
2012	5	64	1546	240	12.80
2011	6	87	2115	239	14.50
2010	8	100	2605	202	12.50
2009	6	89	2208	145	14.83
2008	6	88	2190	114	14.67
2007	4	54	1168	61	13.50
2006	5	52	1246	53	10.40
2005	3	28	646	25	9.33
2004	4	22	660	16	5.50
2003	3	18	482	8	6.00
2002	4	25	427	20	6.25
2001	3	17	466	12	5.67
total	**57**	**644**	**15759**	**1135**	**11.30**

Fonte: SCIELO, 2013[60]

METODOLOGIA

FASE I – Dados Cadastrais – Resultados (CAP. 4)

TABELA 1 – Periódicos encontrados em mais de uma Coleção do SCIELO

ISSN	COLEÇÃO	FASCÍCULOS	ARTIGOS	CITAÇÕES CONCEDIDAS	CITAÇÕES RECEBIDAS
0036-3634	México	98	1.014	35.064	4.761
	Saúde Pública	127	1.279	43.387	2.875
0124-0064	Colômbia	57	644	15.759	1.135
	Saúde Pública	47	575	13.949	302
0213-9111	Espanha	71	866	24.963	3.411
	Saúde Pública	68	826	22.552	2.019
0864-3466	Cuba	60	666	12.279	1.610
	Saúde Pública	25	334	7.147	347
1135-5727	Espanha	97	811	25.235	2.040
	Saúde Pública	96	805	23.697	971
1413-8123	Brasil	91	2.435	75.377	7.951
	Saúde Pública	90	2.405	74.338	2.618
0102-311X	Brasil	220	3.990	116.679	28.050
	Saúde Pública	220	3.991	116.703	13.581
0034-8910	Brasil	263	3.511	67.724	25.292
	Saúde Pública	263	3.511	67.720	11.756
1415-790X	Brasil	57	672	19.600	2.861
	Saúde Pública	57	672	19.600	1.392

Fonte: desenvolvida pelo autor

METODOLOGIA

FASE I – Dados Cadastrais – Resultados (CAP. 4)

TABELA 4 – Dados fontes do SCIELO: diferenças para o ISSN 0213-9111

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Espanha	CSP	Espanha	CSP	Espanha	CSP	Espanha
2002	6	6	58	58	1.509	1.507	98	107
2003	7	7	72	73	1.838	1.843	110	128
2004	7	8	93	102	2.609	2.784	95	188
2005	6	6	58	58	1.527	1.527	133	172
2006	6	6	63	63	128	1.521	20	267
2007	6	6	67	67	1.694	1.639	137	182
2008	6	6	79	79	2.339	2.339	250	352
2009	6	6	93	93	3.136	3.136	304	447
2010	6	6	78	78	2.383	2.383	293	453
2011	6	6	80	80	2.638	2.678	298	516
2012	6	6	85	85	2.751	2.751	281	490
2013	-	2	-	30	-	855	-	109
	68	71	826	866	22.552	24.963	2.019	3.411

Fonte: desenvolvida pelo autor

METODOLOGIA

FASE I – Dados Cadastrais – Análise (CAP. 7)

TABELA 14 – Resumo das diferenças entre o total de citações concedidas: dados fontes do SCIELO

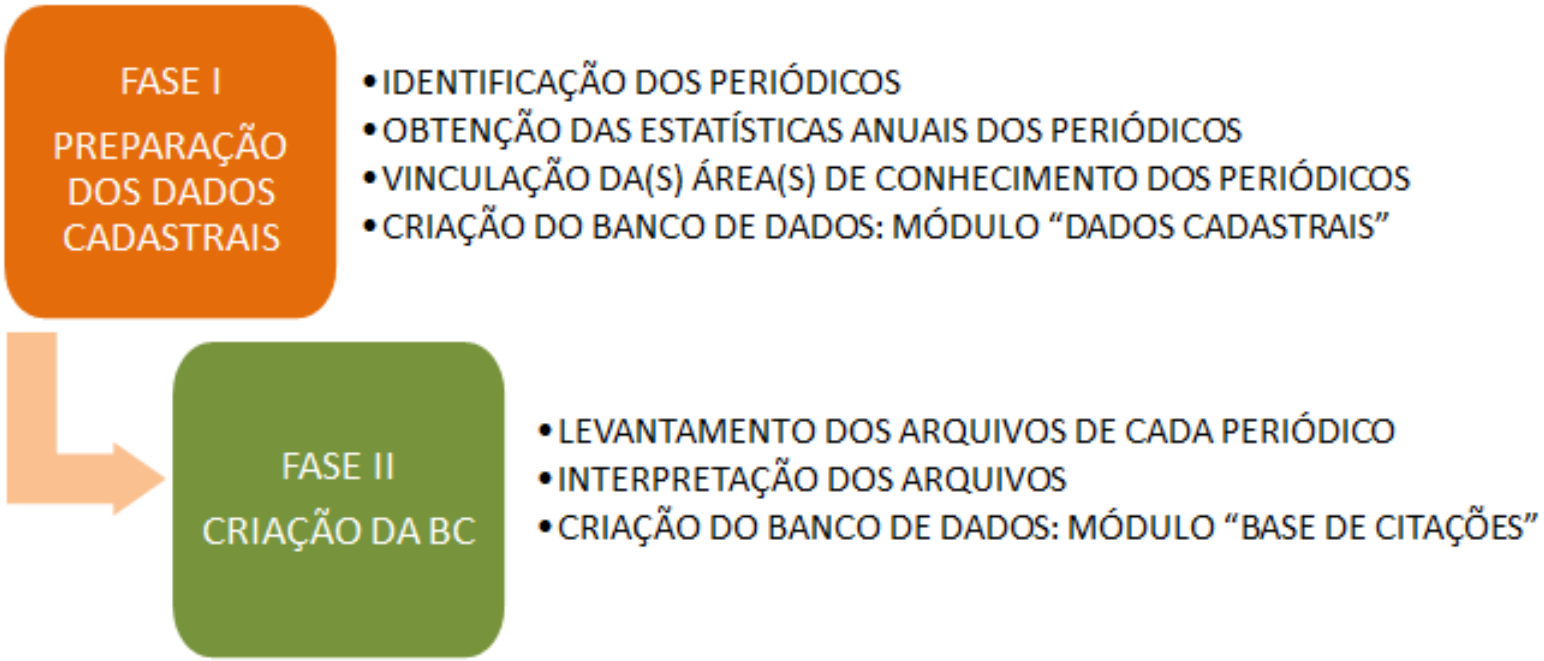
	0036-3634	0124-0064	0213-9111	0864-3466	1135-5727	1413-8123	0102-311x	0034-8910
1.995				100,0%				
1.996				100,0%				
1.997	-100,0%			100,0%				
1.998	-100,0%			100,0%				
1.999	-100,0%			100,0%				
2.000	-100,0%			100,0%				
2.001		100,0%		100,0%				
2.002		100,0%	-0,1%	100,0%				
2.003		75,5%	0,3%	100,0%	1,4%		-1,1%	
2.004			6,3%	100,0%				
2.005				100,0%				
2.006		19,5%	91,6%	100,0%	96,7%			
2.007		-0,3%	-3,4%	-46,2%	-6,5%		0,4%	
2.008	0,1%			-5,7%	-0,9%			
2.009				-15,8%				
2.010		-0,8%						
2.011		0,8%	1,5%	26,7%				
2.012	-14,1%	20,6%						0,1%
2.013	-100,0%		100,0%	100,0%	100,0%	18,4%		

Fonte: desenvolvida pelo autor

METODOLOGIA

FASE II – Base de Citações

FIGURA 2 – Fases para a criação da base de citações do SciELO



Fonte: Desenvolvida pelo autor

METODOLOGIA

FASE II – Base de Citações

FIGURA 5 – Processo automático e contínuo de obtenção e interpretação dos arquivos XML



Fonte: Desenvolvida pelo autor

METODOLOGIA



- Metadados e detalhes de cada artigo

METODOLOGIA

FASE II – Base de Citações

http://www.scielosp.org/scieloOrg/php/ar

onde o IDENTIFICADOR correspond

Assim, S0102-311X20130007000

ISSN = 0102-311X

ANO = 2013

NUMERO = 0007

Sequencial dentro do núm

E o acesso para o arquivo é poss

http://www.scielosp.org/scieloOrg/php/article)



Acta Scientiarum. Agronomy
ISSN 1807-8621

Data do último processamento 29-07-2013

Lista de dados fonte

▽ - clique para selecionar a coluna de ordenação
△ - indica a ordem corrente

título da revista/ano △	n. de fascículos ▽	n. de artigos ▽	n. de citações concedidas ▽	n. de citações recebidas ▽	média de artigos por fascículo ▽
Acta Sci., Agron	**20**	**430**	**10599**	**1438**	**21.50**
2013	2	30	1002	165	15.00
2012	4	60	1476	276	15.00
2011	4	100	2452	372	25.00
2010	4	100	2311	302	25.00
2009	4	100	2307	239	25.00
2008	2	40	1051	84	20.00
total	**20**	**430**	**10599**	**1438**	**21.50**

- Dados fonte de cada periódico (scielo_estatistica – 7.324 registros)
  - ISSN, ano, fascículos, artigos

METODOLOGIA

FASE II – Base de Citações

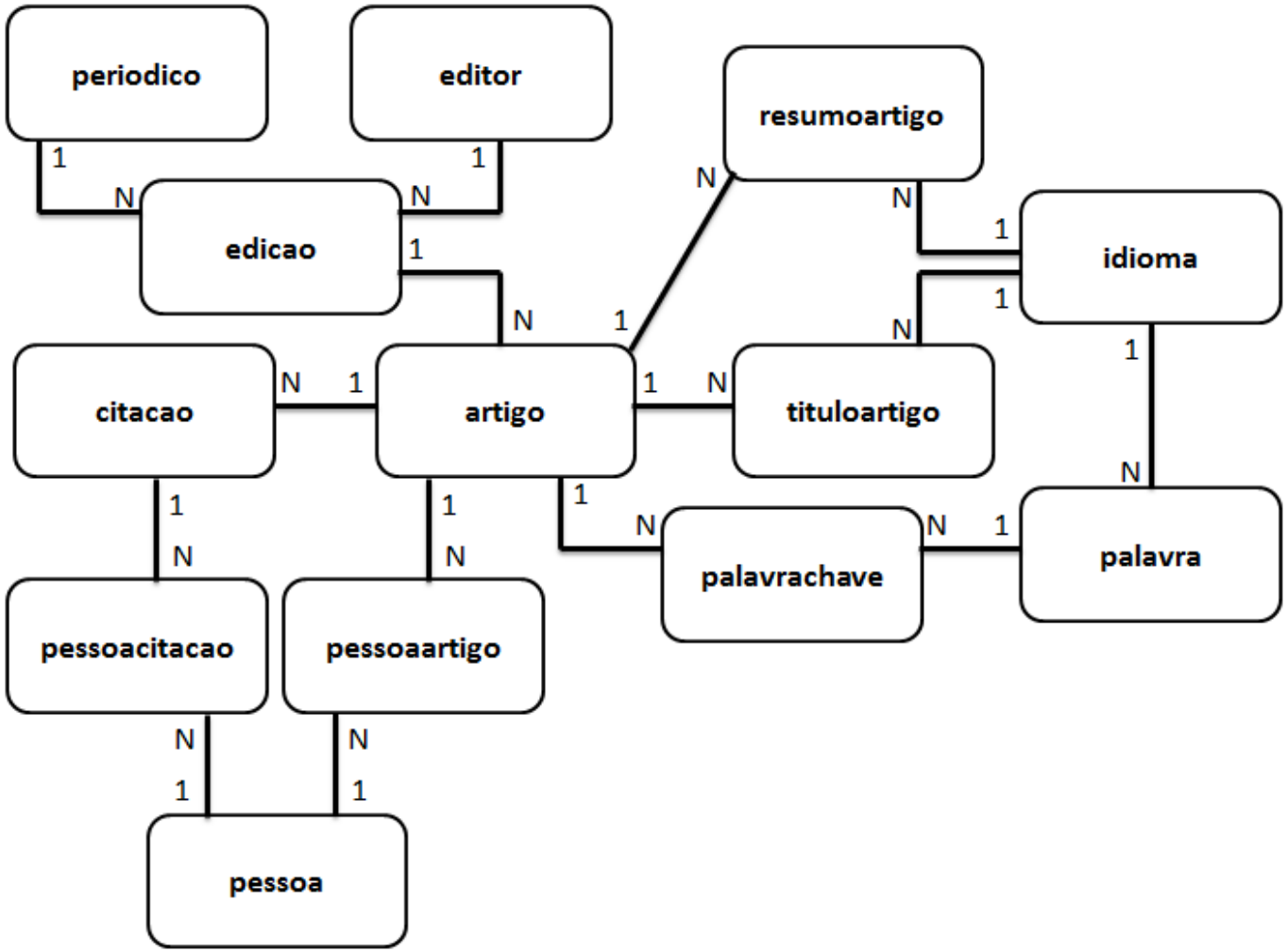


FIGURA 18 – Módulo "Base de Citações"

Fonte: Desenvolvida pelo autor

RESULTADOS DA FASE II - PCI

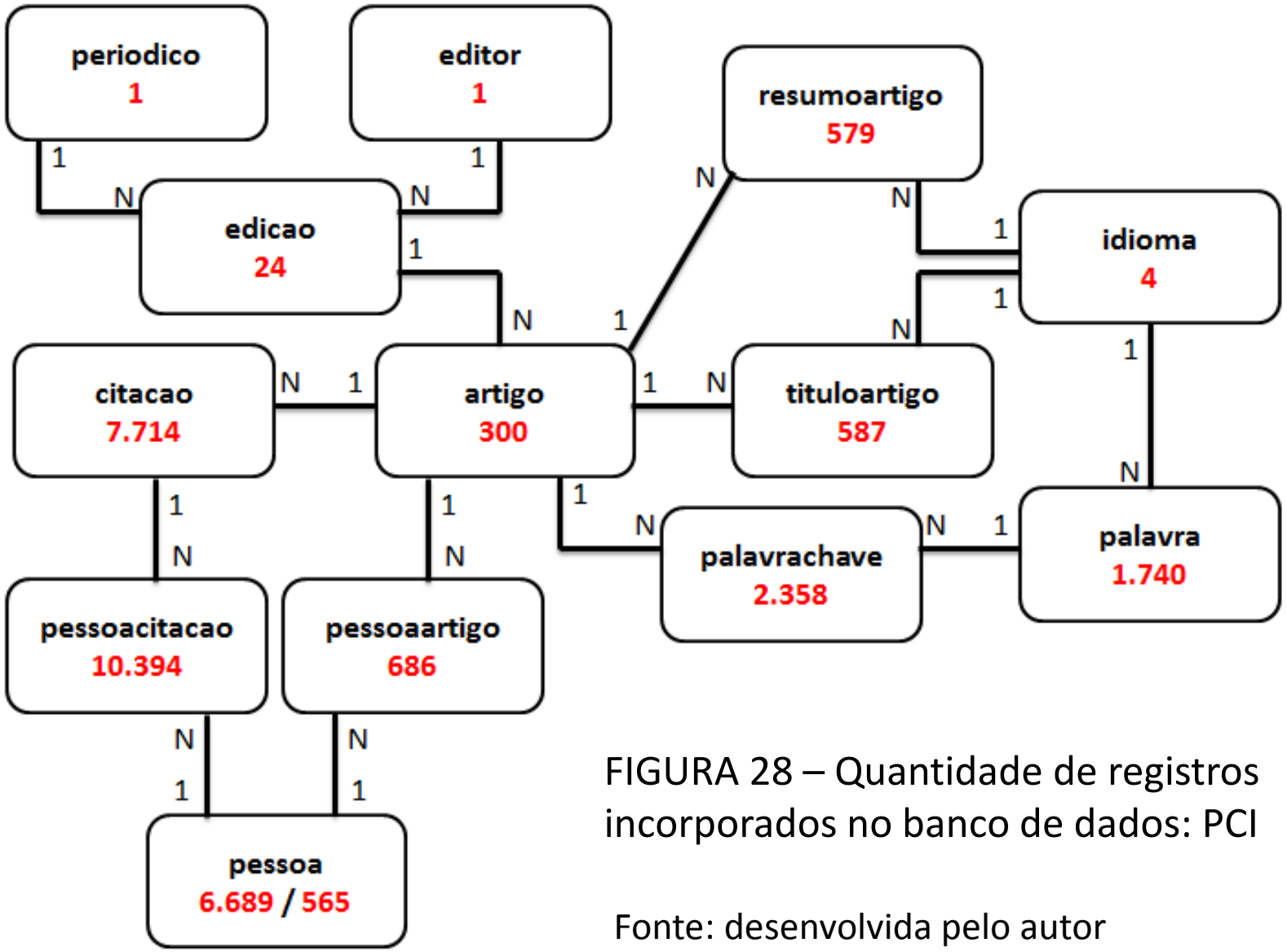


FIGURA 28 – Quantidade de registros incorporados no banco de dados: PCI

Fonte: desenvolvida pelo autor

METODOLOGIA

ISSN	COLEÇÃO	FASCÍCULOS	ARTIGOS	CITAÇÕES CONCEDIDAS
	México	98	1.014	35.064
0036-3634	Saúde Pública	127	1.279	43.387
	XML	80	1.131	34.792
	Colômbia	57	644	15.759
0124-0064	Saúde Pública	47	575	13.949
	XML	54	643	15.726
	Espanha	71	866	24.963
0213-9111	Saúde Pública	68	826	22.552
	XML	69	1.006	24.235
	Cuba	60	666	12.279
0864-3466	Saúde Pública	25	334	7.147
	XML	57	612	11.886
	Espanha	97	811	25.235
1135-5727	Saúde Pública	96	805	23.697
	XML	96	835	24.944
	Brasil	91	2.435	75.377
1413-8123	Saúde Pública	90	2.405	74.338
	XML	87	2.546	74.338
	Brasil	220	3.990	116.679
0102-311X	Saúde Pública	220	3.991	116.703
	XML	199	4.053	116.024
	Brasil	263	3.511	67.724
0034-8910	Saúde Pública	263	3.511	67.720
	XML	253	3.473	67.157
	Brasil	57	672	19.600
1415-790X	Saúde Pública	57	672	19.600
	XML	56	670	19.107

FASE II – Base de Citações – Análise (CAP. 7)

FASE I x FASE II

TABELA 16 – Periódicos encontrados em mais de uma Coleção do SciELO e dados XML

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - PCI

37’

5 RESULTADOS DA FASE II: BASE DE CITAÇÕES DA PERSPECTIVAS EM CIÊNCIA DA INFORMAÇÃO 75

5.1 Obtenção dos dados do SCIELO e criação da Base de Citações 75

5.2 Descrição da amostra: a PCI em números 83

RESULTADOS DA FASE II - PCI

FIGURA 22 – Resumo da importação de dados do SciELO: dados fonte e arquivos XML

1413-9936	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA
TOTAL	**24**	**24**	**0**	**292**	**290**	**2 (0,69%)**	**12,17**	**12,08**	**0,09 (0,74%)**	**7.714**	**7.714**	**0**
2013	1	1	0	11	10	1 (10,00%)	11,00	10,00	1,00 (10,00%)	289	289	0
2012	3	3	0	34	34	0	11,33	11,33	0,00	1.038	1.038	0
2011	4	4	0	50	50	0	12,50	12,50	0,00	1.365	1.365	0
2010	3	3	0	40	39	1 (2,56%)	13,33	13,00	0,33 (2,54%)	1.192	1.192	0
2009	4	4	0	55	55	0	13,75	13,75	0,00	1.353	1.353	0
2008	3	3	0	39	39	0	13,00	13,00	0,00	1.050	1.050	0
2007	3	3	0	35	35	0	11,67	11,67	0,00	873	873	0
2006	3	3	0	28	28	0	9,33	9,33	0,00	554	554	0

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - PCI

FIGURA 29 – Frequências de citações: PCI

FONTE CITADA	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Ciência da Informação	273	2	42	21	38	52	45	40	33	-	-	-
Perspectivas em Ciência da Informação	108	2	10	21	30	13	19	7	6	-	-	-
Scientometrics	71	-	15	6	13	-	23	9	5	-	-	-
Transinformação	55	2	11	11	12	5	3	4	7	-	-	-
Journal of Documentation	47	-	5	8	12	5	8	7	2	-	-	-
DataGramaZero	45	3	16	8	3	7	4	1	3	-	-	-
Journal of the American Society for Information Science	45	-	4	6	3	3	6	16	7	-	-	-
Journal of Information Science	44	-	4	5	12	8	6	4	5	-	-	-
Journal of knowledge Management	35	2	-	10	16	4	1	2	-	-	-	-
Ci. Inf.	28	1	7	10	-	5	2	3	-	-	-	-

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - PCI

FIGURA 30 – Estatística dos periódicos citados: PCI

FONTE CITADA	CITANTE	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Ciência da Informação		273	2	42	21	38	52	45	40	33	-	-	0
	Perspect. ciênc. inf.	273	2	42	21	38	52	45	40	33	-	-	0
Perspectivas em Ciência da Informação		108	2	10	21	30	13	19	7	6	-	-	0
	Perspect. ciênc. inf.	108	2	10	21	30	13	19	7	6	-	-	0
Scientometrics		71	-	15	6	13	-	23	9	5	-	-	0
	Perspect. ciênc. inf.	71	-	15	6	13	-	23	9	5	-	-	0
Transinformação		55	2	11	11	12	5	3	4	7	-	-	0
	Perspect. ciênc. inf.	55	2	11	11	12	5	3	4	7	-	-	0
Journal of Documentation		47	-	5	8	12	5	8	7	2	-	-	0
	Perspect. ciênc. inf.	47	-	5	8	12	5	8	7	2	-	-	0
DataGramaZero		45	3	16	8	3	7	4	1	3	-	-	0
	Perspect. ciênc. inf.	45	3	16	8	3	7	4	1	3	-	-	0
Journal of the American Society for Information Science		45	-	4	6	3	3	6	16	7	-	-	0
	Perspect. ciênc. inf.	45	-	4	6	3	3	6	16	7	-	-	0
Journal of Information Science		44	-	4	5	12	8	6	4	5	-	-	0
	Perspect. ciênc. inf.	44	-	4	5	12	8	6	4	5	-	-	0
Journal of Knowledge Management		35	2	-	10	16	4	1	2	-	-	-	0
	Perspect. ciênc. inf.	35	2	-	10	16	4	1	2	-	-	-	0
Ci. Inf.		28	1	7	10	-	5	2	3	-	-	-	0
	Perspect. ciênc. inf.	28	1	7	10	-	5	2	3	-	-	-	0

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - PCI

FIGURA 31 – Estatística dos periódicos citantes: PCI

CITANTE	FONTE CITADA	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Perspect. ciênc. inf.		7.714	289	1.038	1.365	1.192	1.353	1.050	873	554	-	-	0
	Ciência da Informação	273	2	42	21	38	52	45	40	33	-	-	0
	Anais...	209	16	24	41	31	31	15	7	44	-	-	0
		116	1	32	18	9	13	25	11	7	-	-	0
	Perspectivas em Ciência da Informação	108	2	10	21	30	13	19	7	6	-	-	0
	Scientometrics	71	-	15	6	13	-	23	9	5	-	-	0
	Anais	56	-	-	3	9	30	-	9	5	-	-	0
	Transinformação	55	2	11	11	12	5	3	4	7	-	-	0
	Journal of Documentation	47	-	5	8	12	5	8	7	2	-	-	0
	DataGramaZero	45	3	16	8	3	7	4	1	3	-	-	0
Journal of the American Society for Information Science		45	-	4	6	3	3	6	16	7	-	-	0
	Outros	6.689	263	879	1.222	1.032	1.194	902	762	435	-	-	0

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - PCI

FIGURA 32 – Autores mais citados: PCI

CITANTE / AUTOR CITADO	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Perspect. ciênc. inf. (1413-9936)	10.394	406	1.227	1.966	1.694	1.784	1.487	1.094	736	-	-	0
NONAKA, I.	69	1	3	25	10	8	13	7	2	-	-	0
CHOO, C. W.	35	1	3	5	6	6	9	5	-	-	-	0
PINHEIRO, L. V. R.	33	1	6	10	7	1	4	2	2	-	-	0
CASTELLS, M.	32	1	2	8	5	3	6	6	1	-	-	0
DAVENPORT, T. H.	29	1	4	3	4	7	2	6	2	-	-	0
TAKEUCHI, H.	29	1	1	3	8	3	6	4	3	-	-	0
CAPURRO, R.	28	-	4	5	4	-	5	10	-	-	-	0
GONZÁLEZ DE GÓMEZ, M. N.	28	-	2	4	9	4	8	-	1	-	-	0
SARACEVIC, T.	27	-	3	5	7	1	7	1	3	-	-	0
LÉVY, P.	26	1	3	6	5	3	4	2	2	-	-	0
Outros	10.058	399	1.196	1.892	1.629	1.748	1.423	1.051	720	-	-	0

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - PCI

FIGURA 33 – Palavras-chave mais utilizadas: PCI

CITANTE / PALAVRA-CHAVE	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Perspect. ciênc. inf. (1413-9936)	2.358	78	287	396	335	452	332	259	219	-	-	0
Ciência da informação (Português)	25	-	3	5	4	4	3	2	4	-	-	0
Information Science (Inglês)	22	-	3	5	2	4	2	2	4	-	-	0
Gestão do conhecimento (Português)	19	1	1	1	4	4	2	4	2	-	-	0
Knowledge management (Inglês)	19	1	1	2	4	3	2	4	2	-	-	0
Bibliometria (Português)	15	1	3	1	2	1	2	4	1	-	-	0
Informação (Português)	14	-	-	3	-	4	3	1	3	-	-	0
information (Inglês)	14	-	-	3	-	4	3	1	3	-	-	0
Produção científica (Português)	13	1	2	2	1	2	2	2	1	-	-	0
Scientific production (Inglês)	12	1	2	2	1	2	2	2	-	-	-	0
Gestão da Informação (Português)	10	-	-	1	2	1	2	4	-	-	-	0
Outros	2.195	73	272	371	315	423	309	233	199	-	-	0

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - PCI

FIGURA 34 – Autores que mais produziram: PCI

CITANTE	AUTOR PRODUTOR	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Perspect. ciênc. inf. (1413-9936)		686	38	72	124	89	114	94	79	76	-	-	0
	Bufrem, Leilah Santiago	6	-	-	2	2	1	-	1	-	-	-	0
	Araujo, Carlos Alberto Ávila	4	-	-	1	1	-	1	1	-	-	-	0
	Boccato, Vera Regina Casari	4	-	-	-	2	1	-	-	1	-	-	0
	CUNHA, Murilo Bastos da	4	-	1	1	-	-	1	1	-	-	-	0
	Fujita, Mariângela Spotti Lopes	4	-	1	-	1	-	-	-	2	-	-	0
	Nagano, Marcelo Seido	4	-	1	1	-	1	1	-	-	-	-	0
	Souza, Renato Rocha	4	1	-	-	1	-	1	-	1	-	-	0
	Todesco, José Leomar	4	1	-	1	2	-	-	-	-	-	-	0
	ARAÚJO, Eliany Alvarenga de	3	-	-	1	-	-	-	1	1	-	-	0
	Arboit, Aline Elis	3	-	1	1	1	-	-	-	-	-	-	0
	Outros	646	36	68	116	79	111	90	75	71	-	-	0

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - CSP

42’

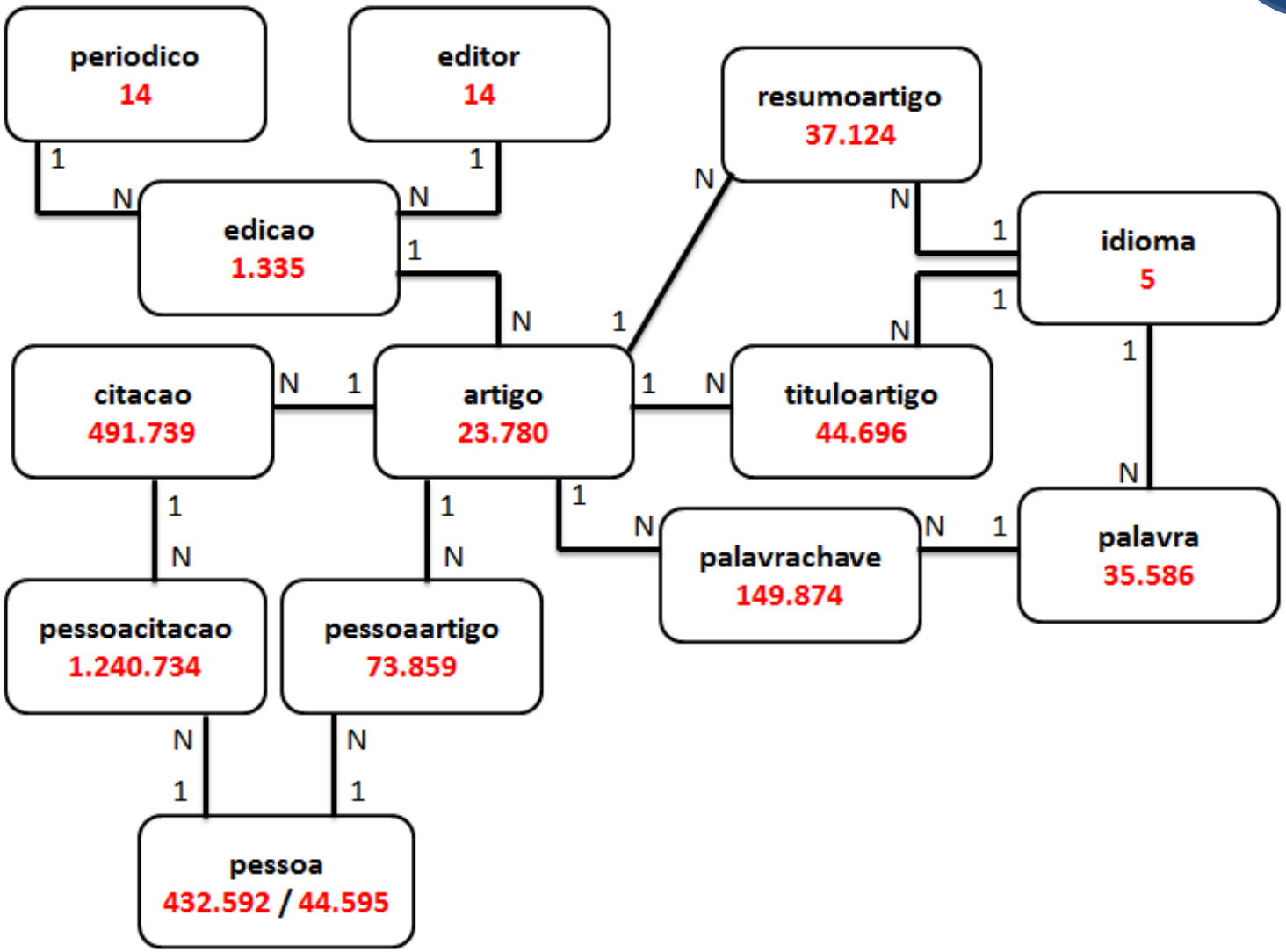
6 RESULTADOS DA FASE II: BASE DE CITAÇÕES DA COLEÇÃO SAÚDE PÚBLICA 93

6.1 Obtenção dos dados do SCIELO e criação da Base de Citações 93

6.2 Descrição da amostra: a Coleção Saúde Pública em números 100

RESULTADOS DA FASE II - CSP

FIGURA 41 – Quantidade de registros incorporados no banco de dados: CSPública



Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - CSP

FIGURA 35

Resumo dos dados de importação das citações da Coleção de Saúde Pública do SciELO

	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
ISSN	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA
TOTAL (14)	**1.330**	**1.347**	**-17 (-1,26%)**	**18.689**	**17.672**	**1.017 (5,75%)**	**14,05**	**13,12**	**0,93 (7,09%)**	**491.613**	**488.252**	**3.361 (0,69%)**
1851-8265	26	26	0	200	165	35 (21,21%)	7,69	6,35	1,34 (21,10%)	5.820	5.833	-13 (-0,22%)
1726-4634	12	12	0	311	278	33 (11,87%)	25,92	23,17	2,75 (11,87%)	6.509	6.509	0
1555-7960	6	6	0	55	46	9 (19,57%)	9,17	7,67	1,50 (19,56%)	1.554	1.560	-6 (-0,38%)
1415-790x	55	56	-1 (-1,79%)	666	651	15 (2,30%)	12,11	11,63	0,48 (4,13%)	18.983	18.997	-14 (-0,07%)
1413-8123	84	87	-3 (-3,45%)	2.442	2.315	127 (5,49%)	29,07	26,61	2,46 (9,24%)	71.386	71.512	-126 (-0,18%)
1135-5727	96	96	0	835	805	30 (3,73%)	8,70	8,39	0,31 (3,69%)	24.944	23.697	1.247 (5,26%)
1020-4989	184	186	-2 (-1,08%)	1.674	1.769	-95 (-5,37%)	9,10	9,51	-0,41 (-4,31%)	48.705	49.858	-1.153 (-2,31%)
0864-3466	56	25	31 (124,00%)	610	334	276 (82,63%)	10,89	13,36	-2,47 (-18,49%)	11.879	7.147	4.732 (66,21%)
0213-9111	69	70	-1 (-1,43%)	1.006	851	155 (18,21%)	14,58	12,16	2,42 (19,90%)	24.235	24.534	-299 (-1,22%)
0124-0064	53	47	6 (12,77%)	628	575	53 (9,22%)	11,85	12,23	-0,38 (-3,11%)	15.391	13.949	1.442 (10,34%)
0102-311x	197	218	-21 (-9,63%)	4.001	3.955	46 (1,16%)	20,31	18,14	2,17 (11,96%)	114.408	115.388	-980 (-0,85%)
0042-9686	159	159	0	1.702	1.428	274 (19,19%)	10,70	8,98	1,72 (19,15%)	46.674	47.047	-373 (-0,79%)
0036-3634	80	98	-18 (-18,37%)	1.131	1.014	117 (11,54%)	14,14	10,35	3,79 (36,62%)	34.792	35.064	-272 (-0,78%)
0034-8910	253	261	-8 (-3,07%)	3.428	3.486	-58 (-1,66%)	13,55	13,36	0,19 (1,42%)	66.333	67.157	-824 (-1,23%)

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - CSP

FIGURA 37

Resumo da importação de dados do SciELO: dados fonte X arquivos XML: ISSN 0124-0064

0124-0064	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA
TOTAL	**53**	**47**	**6 (12,77%)**	**628**	**575**	**53 (9,22%)**	**11,85**	**12,23**	**-0,38 (-3,11%)**	**15.391**	**13.949**	**1.442 (10,34%)**
2012	3	4	-1 (-25,00%)	50	50	0	16,67	12,50	4,17 (33,36%)	1.228	1.228	0
2011	6	6	0	86	87	-1 (-1,15%)	14,33	14,50	-0,17 (-1,17%)	2.082	2.099	-17 (-0,81%)
2010	7	8	-1 (-12,50%)	101	100	1 (1,00%)	14,43	12,50	1,93 (15,44%)	2.605	2.625	-20 (-0,76%)
2009	6	6	0	89	89	0	14,83	14,83	0,00	2.208	2.208	0
2008	6	6	0	88	88	0	14,67	14,67	0,00	2.190	2.190	0
2007	4	4	0	54	54	0	13,50	13,50	0,00	1.168	1.172	-4 (-0,34%)
2006	4	5	-1 (-20,00%)	52	52	0	13,00	10,40	2,60 (25,00%)	1.246	1.003	243 (24,23%)
2005	3	3	0	28	28	0	9,33	9,33	0,00	646	646	0
2004	4	4	0	22	22	0	5,50	5,50	0,00	660	660	0
2003	3	1	2 (200,00%)	18	5	13 (260,00%)	6,00	5,00	1,00 (20,00%)	482	118	364 (308,47%)
2002	4	0	4	23	0	23	5,75	0,00	5,75	410	0	410
2001	3	0	3	17	0	17	5,67	0,00	5,67	466	0	466

Fonte: desenvolvida pelo autor

RESULTADOS DA FASE II - CSP

Problemas identificados que podem justificar diferenças

- Exclusão de informações do SciELO

0124-0064	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA
2002	4	0	4	23	0	23	5,75	0,00	5,75	410	0	410
2001	3	0	3	17	0	17	5,67	0,00	5,67	466	0	466

- Estrutura incompleta de TAGS XML

FIGURA 38 – Falha na estrutura das tags XML dos arquivos interpretados: ISSN 0124-0064

SITUAÇÃO - 0124-0064	ESTRUTURA DOS ARTIGOS NA BC (tags XML)	ARTIGOS BC
Artigos sem citações incorporadas	article; front; /front; back; ref-list; ref-id; /ref; Referências identificadas: 17/17 Ano: 2011	1

Fonte: desenvolvida pelo autor[33]

- Indisponibilidade de arquivos XML

ANÁLISE

46’

7. ANÁLISE DOS RESULTADOS 109
7.1 Análise da obtenção dos dados do SCIELO 109
7.2 Análise da criação da base de citações 116
7.3 Falta de padronização de nomes de autores e fontes 125

ANÁLISE

FASE II – Base de Citações

- Importação foi considerada satisfatória: > 260.000 arquivos XML
- As 4 situações identificadas, caso corrigidas, podem reduzir significativamente as diferenças encontradas
- Sugere-se uma análise mais aprofundada da metodologia SciELO
- Falta de padronização de nomes de autores e fontes

CONCLUSÕES E RECOMENDAÇÕES

48’

8. CONCLUSÕES E RECOMENDAÇÕES 131
8.1 Próximos passos 131
8.2 Futuras pesquisas 131
8.2.1 Desambiguação automática de nomes de autores e fontes 131
8.2.2 Avaliação automática de coleções digitais: uma proposta para o PORTAL CAPES 133
8.3 Considerações finais: a relevância do trabalho realizado 133

CONCLUSÕES E RECOMENDAÇÕES

Próximos passos

- Criação de bases de citação para cada periódico do SciELO
- Criação de base centralizada com todos os periódicos
  para consultas da produção científica nacional registrada no SciELO
- Gestão de conteúdo: criação de interfaces para desambiguação manual
  Vinculação com a graduação da ECI-UFMG

CONCLUSÕES E RECOMENDAÇÕES

Desambiguação automática de nomes de autores e de fontes

- Muita informação incompatível e redundante em SICT (LATTES, OJS/SEER, QUALIS, COLETA, BDTD)
- Vinculação por CPF / Researcher ID / Google Acadêmico
  - Aspectos culturais, políticos, tecnológicos
  - FAPESP: exige Researcher ID
- CPF informado no OJS/evento: dados do LATTES
  - Evita duplicação de nome em eventos por grafia diferente
  - Distingue homônimos
- ISSN: obtenção automática do QUALIS
- Exemplo: Portal de Periódicos OJS

CONCLUSÕES E RECOMENDAÇÕES

Avaliação automática de coleções: PORTAL CAPES

- A partir das fontes armazenadas na base de citações
  - Para determinado periódico ou Coleção
    - Fazer a consulta automática ao PORTAL CAPES
    - Identificar fontes disponíveis no PORTAL…
      … e fontes usadas mas não disponíveis

    - Repetir o processo para outras bibliotecas digitais
      - OJS/SEER

CONCLUSÕES E RECOMENDAÇÕES

Relevância do trabalho

- Possibilidade de…
  - estudos posteriores por área, periódico, coleção…
  - aumento da visibilidade dos periódicos do SciELO
  - automação da extração de dados estatísticos

Criação de uma metodologia que permtirá a interpretação de todos os arquivos XML do SciELO
oferecendo a possibilidade de criação de uma *Web of Science* para a América Latina, Caribe e outras Coleções do SciELO

CONCLUSÕES E RECOMENDAÇÕES

Possibilidades…

FIGURA 20 – Periódicos SciELO agrupados por área de conhecimento

	Periódicos	Fascículos		Artigos		Citações Concedidas		Citações Recebidas	
8 áreas	**904**	**29.728**	**100,00%**	**373.370**	**100,00%**	**9.823.290**	**100,00%**	**729.546**	**100,00%**
CIENCIAS AGRARIAS	96	3.133	10,54%	58.057	15,55%	1.383.257	14,08%	162.645	22,29%
CIENCIAS BIOLOGICAS	94	3.472	11,68%	49.493	13,26%	1.382.040	14,07%	102.499	14,05%
CIENCIAS DA SAUDE	290	12.565	42,27%	151.513	40,58%	3.769.443	38,37%	333.465	45,71%
CIENCIAS EXATAS E DA TERRA	67	2.052	6,90%	28.331	7,59%	818.982	8,34%	31.946	4,38%
CIENCIAS HUMANAS	226	4.232	14,24%	44.425	11,90%	1.332.891	13,57%	61.560	8,44%
CIENCIAS SOCIAIS APLICADAS	148	2.405	8,09%	20.573	5,51%	633.131	6,45%	23.364	3,20%
ENGENHARIAS	70	1.545	5,20%	18.208	4,88%	427.500	4,35%	12.677	1,74%
LINGUISTICA LETRAS E ARTES	23	324	1,09%	2.770	0,74%	76.046	0,77%	1.390	0,19%

Fonte: desenvolvida pelo autor

REVISÃO DOS OBJETIVOS

Objetivos

- ☑ Descrição de uma metodologia para a criação de uma base de citações atualizada automática e continuamente a partir dos periódicos cadastrados no SciELO
  - ☑ Desenvolver um protótipo para testar a metodologia
  - ☑ Validar o protótipo para testar sua aplicação prática
  - ☑ Identificar fatores limitadores e propor alternativas

"Nosso fascínio pela tecnologia nos fez esquecer
o objetivo principal da informação: informar.

Todos os computadores do mundo de nada servirão
se seus usuários não estiverem interessados
na informação que esses computadores podem gerar."

DAVENPORT, 1998

OBRIGADO!