



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-8 – Informação e Tecnologia

<BrCRIS COMO UM SISTEMA DE RECOMENDAÇÃO CIENTÍFICO-TECNOLÓGICA>

<BrCRIS AS A SCIENTIFIC-TECHNOLOGICAL RECOMMENDATION SYSTEM>

Adilson Luiz Pinto (PGCIN/UFSC), Thiago Magela Rodrigues Dias (CEFET/MG e PGCIN/UFSC),
Vivian Santos Silva (IBICT), Josir Cardoso Gomes (IBICT), Washington Luis R. de Carvalho
Segundo (IBICT)

Modalidade: Trabalho Completo

Resumo: Os sistemas de recomendações são aportes para auxiliar usuários em buscas e visualizações de conteúdo com foco em seu interesse e, no contexto científico, visam recomendar periódicos, pesquisadores ou publicações de uma determinada área ou campo de conhecimento. O presente trabalho objetiva relatar a modelagem de sistemas de recomendação no âmbito do projeto BrCRIS que tem a finalidade de construir uma base de dados integrada da produção científica brasileira a partir dos dados de sistemas já existentes como a plataforma Lattes, o Oasis-Br e também de sistemas internacionais. A metodologia é de natureza aplicada por estar reunindo e testando uma série de conteúdos/dados e suas aplicações de uso e a abordagem é tanto quantitativa como qualitativa por estar reunindo e criando um mecanismo de interoperabilidade e de visualização e aplicação de métricas. Como resultado, foram modelados quatro sistemas de recomendação que estão sendo apresentado à comunidade acadêmica sob forma de seminários e palestras com o intuito de colher as primeiras impressões sobre o sistema bem como sugestões para a construção de novas visualizações.

Palavras-Chave: CRIS. Sistemas de Recomendação. Rede de Colaboração Científica.

Abstract: Recommender systems are contributions to assist users in searching and viewing content focused on their interests and, in the scientific context, aim to recommend journals, researchers, or publications from a particular area or field of knowledge. This paper aims to report on the modeling of recommendation systems in the BrCRIS project, which aims to build an integrated database of Brazilian scientific production from data from existing systems such as the Lattes platform, Oasis-Br, and also from international systems. The methodology is based on applied research in that it is gathering and testing a range of content/data and its usage applications, and the approach is both quantitative and qualitative in that it is gathering and creating a mechanism for interoperability and for visualizing and applying metrics. As an initial result, four recommender systems were modeled and are being presented to the academic community in the form of seminars and lectures in order to gather first impressions of the system as well as suggestions for building new visualizations.

Keywords: CRIS. Recommender Systems. Research Networking.

1 INTRODUÇÃO

Os sistemas de recomendações são aportes para auxiliar usuários em buscas de informação e também visualizações de conteúdos com foco em perfil desejado, sendo representado como um filtro de informações colaborativas (CAZELLA et al, 2009), podendo ser um conjunto de dados que, agregados, criam relevância de informação mais próximo de suas necessidades e desejos.

Sistemas de recomendações podem ser criados com opiniões de usuários da comunidade estudada (RESNICK, VARIAN, 1997), mas pode ser criado também por sites comerciais visando identificar novos serviços ou produtos (SCHAFER, KONSTAN, RIEDL, 1999) induzindo os clientes/usuários a preferirem um produto X que outro Y.

Este tipo de sistema de recomendação trabalha com um fator matemático que atrela as respostas dos usuários R, tendo como base a junção de identificar os usuários e os itens a serem avaliados, visando uma pontuação para cada item (R: usuários x itens = pontuação). Assim, seguindo esta lógica chegamos a recomendações dos itens mais buscados (*i*) com pontuações mais expressivas, e também podemos chegar a um conjunto de itens (*k*) em vez de somente um. Este processo de recomendação é destinada para o usuário alvo (*u*) (CAZELLA, 2006).

$$\forall u \in \text{usuários}, i'_u = \arg \text{Max } R(u,i) \\ i \in \text{itens}$$

Para os sistemas que visam recomendar quesitos científicos/acadêmicos o público é mais seletivo e especializado, portanto deve-se atentar que as necessidades e desejos nem sempre serão atendidos com simples perguntas e aplicações de fórmulas pensadas para o mercado de negócios.

Neste sentido alguns modelos já estão sendo aplicados voltados a questões de aprendizagem (BARCELLOS et al., 2000), voltado aos usuários de unidades de informação (KREBS; ROCHA; RIBEIRO, 2017) e para recomendações de trabalhos científicos (GRAVA, 2016; SILVA, 2020).

Para esta proposta iremos desenvolver vários sistemas de recomendações de cunho científico/tecnológico, tendo como base: (1) sistema para reconhecimento de revistas acadêmica de uma determinada área/campo de conhecimento; (2) sistema para reconhecimento de especialistas em determinadas áreas/campos pela sua publicação científica; (3) sistema para reconhecimento de especialistas em determinadas áreas/campos pela sua atividade tecnológica - produção de patentes, e; (4) sistema para reconhecimento

de especialistas em determinadas áreas/campos pelas suas orientações de teses/dissertações e pela participação assídua em bancas de defesas de teses/dissertações.

Esta proposta está baseada na iniciativa BrCris/IBICT, um sistema do tipo *Current Research Information System*¹ (CRIS) que procura armazenar, gerenciar e trocar metadados contextuais para a atividade de pesquisa financiada por órgãos governamentais ou conduzida em uma organização executora de pesquisa. Várias iniciativas para a criação deste tipo de sistema estão sendo desenvolvidas atualmente (BEVAN; HARRINGTON, 2011; JETTEN et al, 2019), com especial destaque para a EuroCRIS.

A EuroCRIS foi criada para agregar instituições com sistemas de informação e em abril de 2021 conta com 329 instituições membros (<https://eurocris.org/community/members>), no qual o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) faz parte desde 2014, denominado como BrCris.

O Projeto BrCris foi inspirado no modelo do Projeto PTCRIS, com natureza de aplicação baseado na interoperabilidade de dados científicos. O modelo português é financiado pela Fundação para Ciência e Tecnologia (FCT), que também é membro da euroCRIS, enquanto a versão brasileira é financiada pelo IBICT e a administração do projeto é executada pelo Fundo de Desenvolvimento da Pesquisa - FUNDEP.

O projeto BrCRIS é um ecossistema de sistemas, com foco em executar a interoperabilidade de dados científicos e tecnológicos. Este sistema integra a troca de dados de vários atores, como agências de financiamento, bases de dados científicas e tecnológicas, repositórios institucionais, plataformas governamentais, infraestrutura de pesquisa, bases de teses e dissertações entre outros atores.

Frente ao universo de possibilidades de sistemas a serem integrados ao projeto BrCris, indagamos de que forma será possível reunir dados/metadados de pesquisa em C&T brasileira em um único espaço? E, ao mesmo tempo, como será sua relação de sistematização destes dados para uma cartografia de forma consistente?

1.1 Objetivos

Pensando na construção de sistemas e na integração para um serviço de recomendações para a sociedade brasileira, a proposta tem como objetivo central a construção e o monitoramento, por interoperabilidade dos dados de ciência e tecnologia

¹ <https://www.eurocris.org/why-does-one-need-cris>

brasileira, em um único sistema, com capilaridades. Estas capilaridades seriam divididas em especificidades, como: (a) Analisar o conjunto de dados recuperados pelos sistemas que o projeto abarca; (b) Tratar, por representações, o conjunto de dados recuperados, visando realizar deduplicação, desambiguação, identificação de entidades, classificação, geolocalização, relações e aplicação de métricas; (c) Disponibilizar um banco de dados com vários segmentos da ciência e da tecnologia normalizados pelo grupo, inclusive com tratamento semântico e com certificação dos dados, e; (d) Representar um sistema de busca e visualização nos quatro níveis de sistemas de recomendações (pesquisadores, revistas, teses/dissertações e patentes).

2 DESENVOLVIMENTO

O desenvolvimento do estudo trabalha com duas vertentes. A primeira procura estudar o que seria ideal em sistemas de recomendações, e a segunda em reunir todo conjunto de dados qualificados.

Partindo da visão geral, temos o que consideramos sistema ideal para sanar informações estratégicas para qualquer tipo de usuário, baseado em Schafer, Konstan e Riedl (2001), porém com uma atenuação do que pretendemos que seja priorizado nos sistemas de recomendações do BrCris.

Figura 1 - Modelo taxonômico do sistema de recomendações do BrCris



Fonte: Extraído de Schafer, Konstan e Riedl (2001, p. 10-11) e adaptado pelos autores.

A segunda parte do sistema de recomendações foi estruturado em três níveis, baseados em Scripts em Python e orquestrando o refinamento dos dados pela Plataforma LA Referencia.

Primeiro nível: A Plataforma LA Referencia foi utilizada para armazenar e exportar os dados coletados, sendo também uma interface de dados de visualização e de busca informacional, servindo para a interoperabilidade de dados com outras plataformas e outros conjuntos de dados. Para o seu desenvolvimento foi criado APIs que permite a relação semântica do universo de dados. A LA Referencia permite que se tenha visibilidade para a produção científica do ensino superior e instituições de pesquisas da América Latina, promovendo acesso livre e gratuito ao texto integral, privilegiando os trabalhos financiados pelo setor público e em acesso aberto.

Segundo nível: Mapeamento de um conjunto de plataformas e bases de dados, como OasisBr/IBICT, BDTD/IBICT, Plataforma Lattes, Espacenet, WikiData, Diadorim/IBICT, Latindex, DOAJ, UlrichsWeb, ROR, Dimensions, Altmetric, Plataforma Sucupira, OpenAIRE, DBpedia, OpenCitations, SherpaRomeo e Google Scholar Metrics. Neste conjunto, algumas plataformas serviram para coleta, outras para homogeneização dos dados e, enfim, outras para certificação dos dados registrados nas demais plataformas;

Terceiro nível: Organização das entidades do conjunto de dados, visando identificar quais são os projetos, serviços, programas de pós-graduação, cursos, organizações e pesquisadores no intuito de obter um arcabouço de dados de publicações, patentes e datasets.

Os principais resultados são sistemas de busca, bem como um painel integrado de visualização de dados (dashboard) que tenha a capacidade de consolidar os dados em uma visão macro sob forma de métricas e indicadores, quantificando assim a presença das organizações científicas e tecnológicas brasileiras e de sua produção.

Para questões metodológicas da proposta, é de natureza aplicada por estar reunindo e testando uma série de conteúdos/dados e aplicações de uso; a abordagem é tanto quantitativa como qualitativa por estar reunindo e criando um mecanismo de interoperabilidade e até mesmo de visualização e aplicação de métricas; em relação aos objetivos, podemos tê-lo como sendo de base exploratória, em função dos testes e até mesmo da criação de um sistema inovador, que acreditamos ser único, na forma em que está construído, no cenário mundial; quanto aos procedimentos caracteriza-se como

experimental por estar testando tecnologias e conjuntos de dados novos, e estar reunindo uma vasta gama de dados heterogênicos e de naturezas distintas.

3 RESULTADOS

A proposta e a modelagem de dados para a construção dos sistemas de recomendações do BrCris seguiu um processo evolutivo, que em 2021 pode ser vislumbrado em uma ferramenta de uso, a princípio para órgãos governamentais, com possibilidade de consulta pública a partir de 2022.

A evolução vem desde 2015 até a data corrente (2021), período em que foi sendo adquirido know how a partir da visão do sistema português e a confirmação de que o IBICT seria a instituição brasileira responsável por tornar real a harmonização e consolidação de vários sistemas em um único local.

A guinada da proposta brasileira passa a ser real quando começam a integrar sistemas, já utilizados em avaliações nacionais, como pacotes da Capes, do CNPq, da SciELO em 2017 e ganha importância quando passa a adotar o modelo de interoperabilidades de dados da LA Referencia em 2019.

Entretanto, o sistema se torna algo voltado a recomendações quando é idealizada a junção de vários produtos nacionais e internacionais de acesso aberto para a consolidação de uma única base de dados, com todas as características possíveis de interoperabilidade e a articulação ontológica pelo projeto da plataforma VIVO (BÖRNER et al, 2012; HEON et al, 2020).

Para uma capilaridade de todos os dados reunidos pelo BrCris foi fundamental discutir os seus aspectos, como: (a) conjunto de dados reunidos, (b) tratamento dos dados a este conjunto, (c) constituição da reunião dos dados em um modelo de dados estruturado para as necessidades da proposta do BrCris, e (d) formato da disponibilização dos dados para à Sociedade.

3.1 Análise do conjunto de dados

Os conjuntos de dados reunidos, especificamente a partir de 2020, teve uma condução na reunião de todos os sistemas livres de informações em ciência e tecnologia voltados ao cenário Brasil, além da idealização da atual equipe do BrCris.

Como base para a normalização dos dados foi utilizada a Plataforma LA Referencia, que foi criada na linguagem de programação Java, conta com um trabalho de mais de sete anos de desenvolvimento e é utilizada com o mesmo propósito por mais de dez países.

Conforme mencionado, esta plataforma trabalha com coleta de dados de repositórios, é utilizada pelo IBICT e outras instituições da América Latina, trabalha com protocolo de coleta de metadados OAI-PMH e serve para validação, transformação e enriquecimento de metadados.

No contexto do BrCris a plataforma trabalha com adição de fontes de informação, no qual processa, identifica entidades e seus relacionamentos de dados, bem como armazena entidades e flexibiliza as ações do ecossistema. A Plataforma também trabalha no sentido de implementar mecanismos para identificar entidades e, de certa forma, gera identificadores únicos, visando fornecer serviços com valor agregado para os atores e agentes do sistema BrCris. Alimenta também outros serviços, como a geração de estatísticas sobre os dados agregados (por meios de métricas e rankings), além de um sistema de busca e a representação dos dados em um conjunto de dashboards.

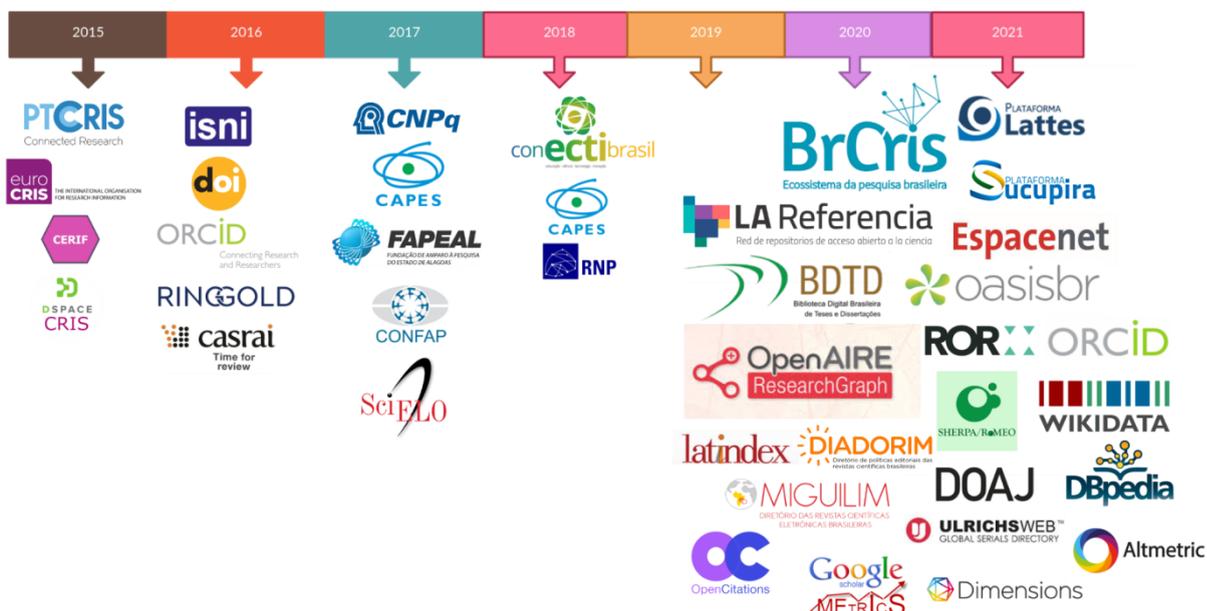
As fontes secundárias de informação para a interoperabilidades dos dados são: Bancos de teses, Bases de dados de conteúdos abertos em ciência, Repositórios de dados, Catálogos científicos, Processos e infraestruturas para pesquisa em acesso aberto, Base de patentes, Sistemas de métricas, Sistemas de Currículos virtuais, Sistemas de classificação de produção científica da pós-graduação e Dados enciclopédicos estruturados por revistas científicas.

Ao final do ano de 2014 houve um constante trabalho por parte do IBICT e sob orientação da equipe do PTCRIS para estudo e adoção de padrões sobre descrição de recursos e interoperabilidade de sistemas de informação científica, identificadores persistentes e realização de parcerias entre instituições do setor público que gerenciavam grande volume de dados sobre o ecossistema da pesquisa científica brasileira.

Somente no ano de 2019 implementou-se um projeto formal com recursos obtidos de emenda parlamentar para o mapeio e organização da informação científica relativa à pesquisa brasileira, que estava dispersa em diferentes bases abertas nacionais e internacionais. Entre estas bases, destacam-se: (1) BDTD e OasisBr; (2) Plataformas Lattes e Sucupira; (3) Plataforma LA Referencia; (4) OpenAIRE Research Gate; (5) Wikidata, e; (6) OpenCitations. As bases mapeadas foram coletadas por meio de acesso à API's públicas e

acesso direto à informação disponível na homepage de cada fonte. Toda a informação coletada foi normalizada de modo a seguir o padrão de importação para a plataforma LA Referencia.

Figura 2 - Histórico de construção do modelo do Sistema BrCRIS



Fonte: Elaborado pelos autores.

3.2 Tratamento dos dados

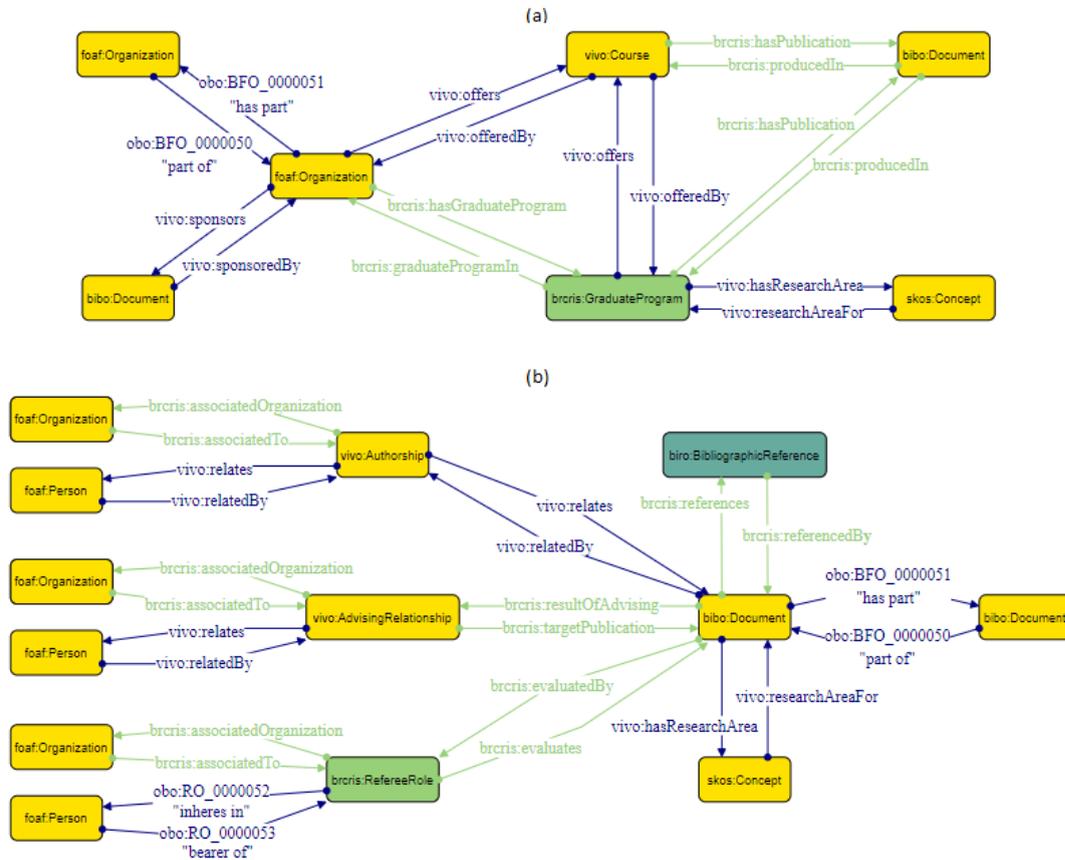
Na questão de obtenção, transformação e tratamento dos dados foi utilizado um extrator incremental que trabalhou com uma listagem de identificadores utilizando o protocolo SOAP para acessar e extrair os currículos da Plataforma Lattes no formato XML. Também foi criado um documento descritivo com todas as entidades possíveis e relações dentro dos conjuntos de dados recuperados de Plataformas e bases de dados.

Outra criação de arquitetura foi estruturar os dados a partir do modelo da Plataforma LA Referencia, adequando os dados do Lattes e de outras fontes indicando quais campos, entidades e relacionamentos deveriam ser armazenadas. Seqüencialmente, foram criadas as entidades de trabalho para o BrCris de acordo com as especificações de cada base extraída, independentemente do formato provido pela fonte. Este processo teve o intuito de validar as definições das novas entidades e a geração dos identificadores únicos BrCris.

Uma etapa seguinte foi a criação dos relacionamentos segundo o modelo do LA Referencia e por fim a exportação destes dados a partir do formato exigido para cada sistema de recomendação. Foram utilizadas as seguinte entidades mapeadas: (1) Person, (2) OrgUnit, (3) 1. Journal; 2. Event, (4) Service, (5) Publication, (6) Dataset, (7) Course, (8)

GraduateProgram, (9) Patent, (10) Project, (11) Funding, (12) ResearchArea, (13) BibliographicReference, (14) ResearchGroupMembership, (15) ProjectMembership e (16) 1. Authorship; 2. Advisoring; 3. Refereeing; 4. Editing.

Figura 3 - Recortes do modelo semântico para (a) Organização e (b) Publicação



Fonte: Elaborado pelos autores.

Para a disponibilização dos dados em formato padronizado e interoperável, o modelo interno de entidades e relacionamentos foi mapeado para um modelo semântico que utiliza como base a ontologia VIVO (BÖRNER, 2012). Esta ontologia descreve o domínio de pesquisa acadêmica, sendo composta por classes e propriedades que representam uma rede de pesquisadores e instituições de pesquisa, os projetos nos quais estão envolvidos e os resultados de suas pesquisas, como publicações, patentes, software, etc. O principal objetivo da ontologia VIVO é permitir a avaliação do impacto de pessoas e organizações, e a identificação e reuso dos produtos de suas atividades de pesquisa, em linha com a finalidade central de geração de métricas do BrCris.

O modelo semântico resultante é, portanto, composto por um subconjunto da ontologia VIVO acrescido de uma extensão local destinada a representar classes e propriedades específicas do contexto de pesquisa acadêmica brasileiro. As classes que

compõem o modelo são: Person, Organization, Service, Journal, Document, Course, Patent, Project, Grant, Event, BibliographicReference, MemberRole, Authorship, AdvisingRelationship, Editorship e Concept - usada para representar Research Areas -, provenientes da ontologia VIVO; e GraduateProgram e RefereeRole, criadas na extensão local. Além de ser um recurso formal que valoriza o reuso de representações já existentes, atendendo a padrões internacionais, este modelo semântico permite a visualização e exploração dos dados em formato de grafo na plataforma VIVO, também bastante utilizada por outros CRIS. A Figura 4 apresenta dois recortes do modelo semântico, centrados nos conceitos de (a) Organização (foaf:Organization) e (b) Publicação (bibo:Document), com as principais classes associadas e as propriedades (diretas e inversas) que estabelecem relacionamentos entre elas.

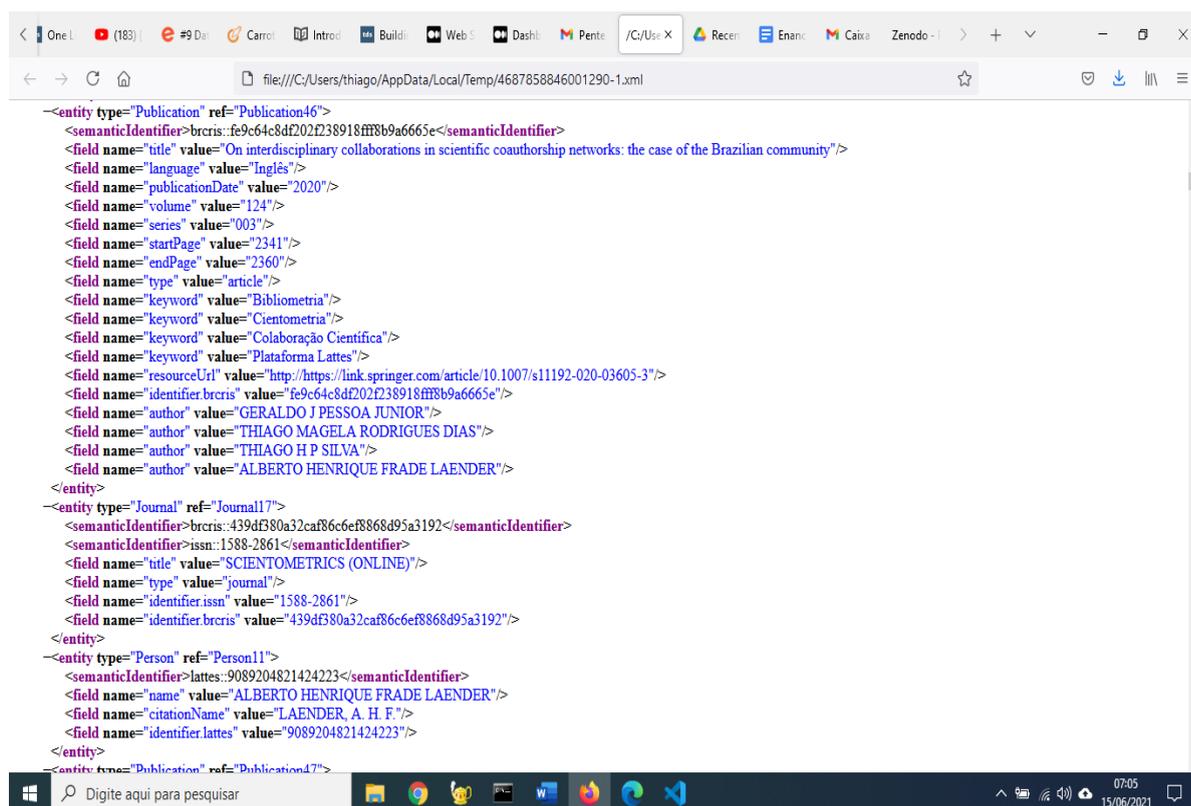
3.3 Conjunto de dados

Entre os serviços previstos sobre a base de dados agregada e organizada está a exportação em formato intercambiável com o DSpace⁷ de publicações, teses, dissertações e patentes, de forma que estas exportações possam ser utilizadas na criação de novos repositórios institucionais e na migração de repositórios baseados na versão anteriores do DSpace, que segue um novo paradigma com relação ao modelo de dados. Do mesmo modo as API's geradas pelo BrCris poderão ser utilizadas nas ferramentas de vocabulário controlado e controle de autoridade no preenchimento dos formulários de submissão dos repositórios institucionais de publicações.

Desse modo, todo o repositório de dados que compõem o BrCris, composto pelas diversas fontes integradas e normalizadas poderá ser utilizado não só para consultas e análises mas também para exportação pelas API's disponibilizadas, em formato padronizado e com conteúdo certificado. Por exemplo, um conjunto de dados que poderia ser utilizado tanto para análises bibliométricas como para definir um conjunto de publicações de um determinado conjunto de indivíduos. Na Figura 5, é possível observar um fragmento do conteúdo de um arquivo que pode ser obtido pelas API's do BrCris contendo as publicações de um determinado pesquisador.

Figura 4 - Fragmento dos dados referente ao conjunto de publicações de um pesquisador

² <https://duraspace.org/dspace/dspace-7>



```
<entity type="Publication" ref="Publication46">
  <semanticIdentifier>brcris:fe9c64c8df202f238918ff8b9a6665e</semanticIdentifier>
  <field name="title" value="On interdisciplinary collaborations in scientific coauthorship networks: the case of the Brazilian community"/>
  <field name="language" value="Inglês"/>
  <field name="publicationDate" value="2020"/>
  <field name="volume" value="124"/>
  <field name="series" value="003"/>
  <field name="startPage" value="2341"/>
  <field name="endPage" value="2360"/>
  <field name="type" value="article"/>
  <field name="keyword" value="Bibliometria"/>
  <field name="keyword" value="Cientometria"/>
  <field name="keyword" value="Colaboração Científica"/>
  <field name="keyword" value="Plataforma Lattes"/>
  <field name="resourceUrl" value="http://https://link.springer.com/article/10.1007/s11192-020-03605-3"/>
  <field name="identifer.brcris" value="fe9c64c8df202f238918ff8b9a6665e"/>
  <field name="author" value="GERALDO J PESSOA JUNIOR"/>
  <field name="author" value="THIAGO MAGELA RODRIGUES DIAS"/>
  <field name="author" value="THIAGO H P SILVA"/>
  <field name="author" value="ALBERTO HENRIQUE FRADE LAENDER"/>
</entity>
<entity type="Journal" ref="Journal7">
  <semanticIdentifier>brcris:439df380a32caf86c6ef8868d95a3192</semanticIdentifier>
  <semanticIdentifier>issn:1588-2861</semanticIdentifier>
  <field name="title" value="SCIENTOMETRICS (ONLINE)"/>
  <field name="type" value="journal"/>
  <field name="identifier.issn" value="1588-2861"/>
  <field name="identifer.brcris" value="439df380a32caf86c6ef8868d95a3192"/>
</entity>
<entity type="Person" ref="Person11">
  <semanticIdentifier>lattes:9089204821424223</semanticIdentifier>
  <field name="name" value="ALBERTO HENRIQUE FRADE LAENDER"/>
  <field name="citationName" value="LAENDER, A. H. F."/>
  <field name="identifer.lattes" value="9089204821424223"/>
</entity>
<entity type="Publication" ref="Publication7">
```

Fonte: Elaborado pelos autores

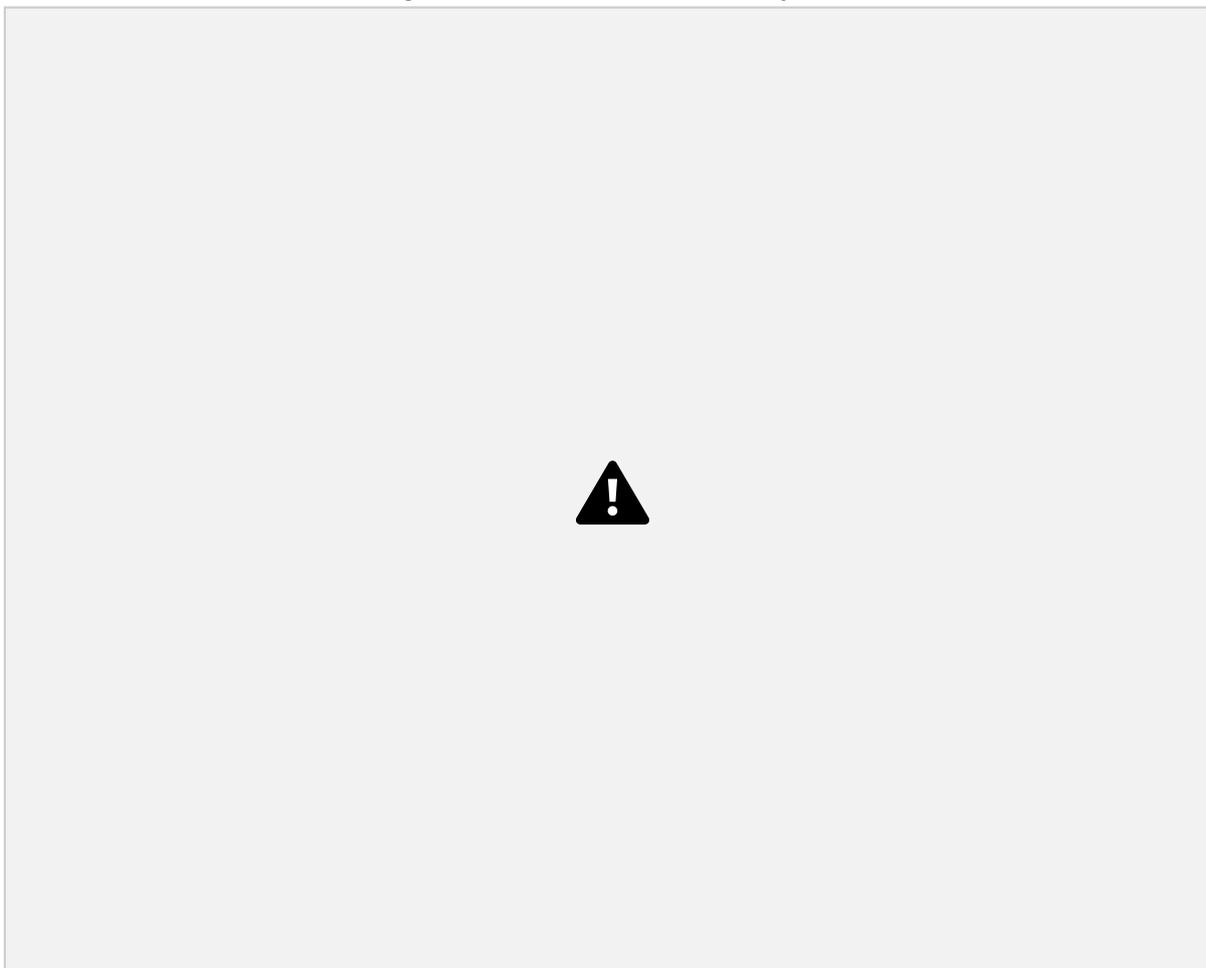
Como pode ser observado, todas as publicações identificadas de um determinado pesquisador são mapeadas para o modelo proposto, em que todos os dados são devidamente classificados e padronizados, viabilizando sua importação por outras ferramentas de análise ou repositórios locais. Os registros agregam informações sobre a publicação em si “Entidade Publication”, inclusive com a geração de um identificador semântico; informações sobre o periódico em que o trabalho foi publicado “Entidade Journal”, e ainda, sobre o conjunto de autores que participaram da publicação “Entidade Person”. Destaca-se que todos os identificadores persistentes, identificados para cada uma das entidades consideradas, são incluídos nos dados da entidade, o que viabiliza o processo de desambiguação dos dados, um grande desafio no processo de integração de grandes repositórios em formatos distintos.

Um conjunto de arquivos contendo dados de publicações no modelo adotado pelo BrCris já está postado no Zenodo para compartilhamento (<https://doi.org/10.5281/zenodo.4955756>).

3.4 Sistemas de recomendações

A possibilidade de se gerar sistemas de recomendações são as mais infinitas possíveis, entretanto para o BrCris apresentamos quatro modelos, com dados complementares e de ações independentes.

Figura 5 - Sistema de recomendações



Fonte: Elaborado pelos autores

O primeira sistema (a - revistas científicas) foi criado um sistema para identificar especialidade de publicações científicas. Os dados utilizados para a sua produção foram bases de dados, como Diadorim, Latindex, Portal ISSN, DOAJ, UlrichsWeb, Plataforma Lattes, Google Scholar Metrics. O foco de análise teve como objetivo verificar: (a) tempo de resposta da submissão até a publicação; (b) h-5 média; (c) h-5 mediana; (d) taxa de publicação; (e) Comitê da revista - nacional ou internacional; (f) endogenia para as revistas brasileiras, e; (g) idiomas de publicação.

O segundo sistema (b - pesquisadores) foi criado para identificar os especialistas de publicações, sejam por produção científica ou até mesmo produções técnicas. Os dados

utilizados foram provenientes da Plataforma Lattes e da normalização de dados pelo La Referencia. O foco de análise foi criado em identificar: (a) índice de publicação; (b) índice de citação; (c) geolocalização de publicação; (d) áreas temáticas, e; (e) índice de coautoria.

O terceiro sistema (c - teses e dissertações) foi criado para identificar os especialistas em orientações e participações de bancas, no qual foi utilizada a base de dados da Biblioteca Digital Brasileira de Teses e Dissertações e a parte de orientações da Plataforma Lattes. O foco de análise foi criado para identificar: (a) Genealogia por até 11 níveis; (b) orientações por regiões; (c) ancestralidade acadêmica; (d) especialistas em temáticas em orientações, e; (e) especialistas em temáticas em participações de bancas.

O quarto sistema (d - patentes) foi criado para identificar especialidades de produtores e inventores de patentes no Brasil, utilizando como suporte a base de dados Espacenet. O foco deste sistema está em analisar: (a) citações recebidas; (b) citações realizadas; (c) concessão e prorrogação; (d) família de patentes; (e) triádica e; (f) classificação.

Todo o processo ainda está na versão beta e deve ser disponível para o público em dezembro/2021, com restrições e com login e senha, justamente para o projeto passa identificar todos os dados de usabilidade dos sistemas.

4 CONSIDERAÇÕES FINAIS

O BrCris está em estágio de conclusão de suas representações, tendo o modelo de dados de agregação já consolidado, sendo que o conjunto de dados recuperados são de bases diversas e de grande capilaridade para pesquisas em Ciência da Informação. Neste sentido, a interoperabilidade dos dados já foi realizada e armazenadas, com deduplicação, na La Referencia, que serve de base para todos os ajustes finos de pesquisadores, instituições, departamentos, revistas, patentes e outros componentes analisados.

A questão de tratamento das representações, todas as ações foram realizadas com sucesso. Até o mês de maio de 2021 foram recuperados os dados das referidas bases e foram identificadas as deduplicação, desambiguação, identificação de entidades, classificação e geolocalização das representações realizadas em VIVO. Outras ações também foram agregadas para a geração de dois dos quatro dashboard utilizando o Kibana e ElasticSearch, foram os casos para identificar revistas e teses/dissertações.

Para a questão das patentes, o processo foi tratado pelo projeto Patent2Net, gerando agregações de clusterização, geolocalização e correlações.

Em relação a disponibilidade dos bancos de dados, já foram feitas as questões semânticas e todos os dados já foram certificados em junho de 2021, porém como são dados extraídos de plataformas interativas o processo de certificação será constante dentro do projeto, justamente para não duplicar esforços de dados ao conjunto que será disponível para a sociedade em breve.

Finalmente, a parte de visualização e constância dos primeiros sistemas de métricas já estão representados no sistema, porém existem muito por fazer na parte de interatividade dos dados tratados, no qual tem-se como plano de ser solucionado até o mês de fevereiro de 2022.

AGRADECIMENTOS

Aos demais participantes do projeto BrCris: Luc Quoniam, Lautaro Matas, Tales Henrique José Moreira, Juliana Schineider, Jesús Mena-Chalco.

REFERÊNCIAS

- BARCELLOS, Carla Duarte et al. Sistema de recomendação acadêmico para apoio a aprendizagem. **Revista Renote**, v. 5, n. 2, p. 1-10, 2007. Disponível em: <https://doi.org/10.22456/1679-1916.14236> . Acesso em: 20 ago. 2021.
- BEVAN, Simon J.; HARRINGTON, John. Managing research publications: lessons learned from the implementation of a current research information system. **Serials: The journal for the Serials Community**, v. 24, n. 1, p. 26-30, 2011. Disponível em: <https://doi.org/10.1629/2426>. Acesso em: 20 ago. 2021.
- BÖRNER, Katy et al. VIVO: A semantic approach to scholarly networking and discovery. **Synthesis Lectures on the Semantic Web: Theory and Technology**, v. 2, n.1, p. 1-178 Indiana: Morgan & Claypool Publishers series, p. 1-178, 2012. Disponível em: <https://doi.org/10.2200/S00428ED1V01Y201207WBE002>. Acesso em: 20 ago. 2021.
- CAZELLA, Silvio Cesar et al. Recomendação de Objetos de Aprendizagem Empregando Filtragem Colaborativa e Competências. In: **Simpósio Brasileiro de Informática na Educação (SBIE)**, Florianópolis, 2009. Disponível em: http://www.niee.ufrgs.br/eventos/SBIE/2009/conteudo/artigos/completos/61791_1.pdf. Acesso em: 20 ago. 2021.
- CAZELLA, Silvio Cesar. **Aplicando a relevância da opinião de usuários em sistema de recomendação para pesquisadores**. Tese (Doutorado) - Programa de Pós-Graduação em Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006. Disponível em: <https://lume.ufrgs.br/handle/10183/8424>. Acesso em: 29 abr. 2021.
- GRAVA, Arthur Patricio. **Sistema de recomendação de artigos científicos utilizando dados sociais**. Dissertação (Mestrado) - Programa de Pós-Graduação em Sistemas de Informação, Universidade de São Paulo, São Paulo, 2016. Disponível em: <https://doi.org/10.11606/D.100.2016.tde-26072016-160726>. Acesso em: 29 abr.2021.

HEON, Michel et al. VIVO as a linked open data enabler for the Université du Québec network. In: **11th International VIVO Conference**, June 16-18, 2020. Disponível em: <https://doi.org/10.13140/RG.2.2.25018.41924>. Acesso em: 19 ago. 2021.

JETTEN, Mijke; SIMONSB, Ed; RIJNDERSC, Jan. The Role of CRIS's in the research life cycle: a case study on implementing a FAIR RDM policy at Radboud University, the Netherlands.

Procedia Computer Science, v. 146, p. 156-165, 2019. Disponível em:

<https://doi.org/10.1016/j.procs.2019.01.090>. Acesso em: 12 ago.2021.

KREBS, Luciana Monteiro; ROCHA, Rafael Port da; RIBEIRO, Cristina. Quem leu este também leu...: sistema de recomendação na biblioteca universitária. **Perspectivas em Ciência da Informação**, v. 22, n. 1, p. 151-169, 2017. Disponível em:

<https://doi.org/10.1590/1981-5344/2496>. Acesso em: 11 maio 2021.

RESNICK, Paul; VARIAN, Hal R. Recommender systems. **Communications of the ACM**, v. 40, n. 3, p. 55-58, 1997. Disponível em: <https://doi.org/10.1145/245108.245121>. Acesso em: 5 jun. 2021.

SCHAFER, J. Ben; KONSTAN, Joseph; RIEDL, John. Recommender systems in e-commerce. In: **ACM Conference on e-Commerce**, 2, 1999. Mineapolis, Proceedings..., New York: ACM Press, p. 158-166, 1999. Disponível em:

https://dl.acm.org/doi/pdf/10.1145/336992.337035?casa_token=6CEavqx_blwAAAAA:P-f1vn9A6tNBO82qiligw3oC3rTiiNVkpDOZCpDpR9FFrOCJdQNmCxrpnXD_t2tWJFJQ_a-PDgoBknM. Acesso em: 11 maio 2021.

SCHAFER, J. Ben; KONSTAN, Joseph; RIEDL, John. E-Commerce recommendation applications. **Data Mining and Knowledge Discovery**, v. 5, n. 2, p. 115-153, 2001. Disponível em:

<https://link.springer.com/content/pdf/10.1023/A:1009804230409.pdf>. Acesso em: 11 maio 2021.

SILVA, Hugo Guilherme Andrade da. **Um estudo de caso de aplicação de sistemas de recomendação baseado em filtragem colaborativa na tarefa de designação de profissionais para serviços de TIC Offshore**. 2020. 89 f. Dissertação (Mestrado Profissional em Engenharia de Produção e Sistemas Computacionais) - Universidade Federal Fluminense, Rio das Ostras, 2020. Disponível em: <https://app.uff.br/riuff/handle/1/21694>. Acesso em: 19 ago. 2021.