

UMA BIBLIOTECA DIGITAL GLOBAL DE MATEMÁTICA

Representando o Committee on Electronic Information and Communication da IMU (CEIC), a professora Olga Caprotti¹ participou do World Data System (WDS) Member's Forum, de caráter bianual, que desta vez ocorreu em Nova Delhi, em 02 de novembro de 2014, para apresentar a ideia da Global Digital Mathematical Library (GDML) para a comunidade de ciência de dados. Ciência de dados é uma área nascente, que estuda a extração de conhecimento a partir de dados, e usa técnicas e teorias de matemática, estatística e tecnologia da informação.

O World Data System (WDS) é um organismo interdisciplinar do International Council for Science (ICSU). O ICSU-WDS se baseia no legado dos World Data Centres (WDC) e da Federation of Astronomical and Geophysical data analysis Services (FAGS) estabelecidos pelo ICSU para lidar com dados gerados pelo Ano Geofísico Internacional (1957–1958).

A International Mathematical Union (IMU) participa como membro associado, e o pôster apresentado foi intitulado "Trusted services and data from the GDML".

Para mais detalhes do Forum, veja www.icsu-wds.org/community/wds-members-forum

A seguir, traduzimos o texto contido no pôster, levando seu conteúdo a ser conhecido no Brasil, por acreditar na importância dessa iniciativa para o futuro próximo da comunidade matemática mundial.

História

A IMU apoia fortemente a implementação de uma Biblioteca Digital Global de Matemática (GDML). Ela tem em conta que os matemáticos podem tanto continuar

utilizando literatura na forma digital de maneiras similares à literatura impressa tradicional, quanto podem aproveitar a tecnologia emergente para avançar o conhecimento.

A Presidente da IMU, Ingrid Daubechies², e Peter Olver, *Chair* do Committee on Electronic Information and Communication da IMU (CEIC), tomaram a iniciativa de trabalhar pela GDML por meio de consultas a um grupo amplo de especialistas. A visão da GDML foi oficializada pela Assembleia Geral da IMU que, já em 2006, havia apoiado um documento, "Digital Mathematics Library: A Vision for the Future", proposto pelo CEIC, dizendo "... apoia a visão de uma coleção distribuída da sabedoria matemática passada que atenda às necessidades de toda a ciência, e encoraja matemáticos e editores de matemática a se unirem para implementar essa visão".

Em 2012, o CEIC organizou o International Symposium on the Future World Heritage Digital Mathematics Library. As conclusões do encontro exploram mecanismos práticos, desafios e capacidades necessárias para a concretização da GDML [1].

Durante o Congresso Internacional de Matemáticos de 2014 (ICM2014), o recentemente formado Grupo de Trabalho GDML, coordenado por Patrick Ion, com suporte da IMU, foi encarregado de esboçar uma estratégia de criação da GDML, determinando sua estrutura organizacional, priorizando requisitos para sua implementação, estimando custos e fomentando a redação de pedidos para agências de financiamento. A discussão, iniciada durante o ICM2014, continua no recém-criado blog da GDML [2].

A IMU apoia o documento Open Access, endossado pelo ICSU na Assembleia Geral de 2014, que contri-

¹ Atualmente Chief Scientist da organização World Education Portals (WEPS).

² N.E.: Ainda presidente quando o pôster foi apresentado, seu mandato foi de 2011 a 2014.

bui com blocos significativos da infraestrutura necessária para a concretização da GDML. John Ball, ex-Chair do CEIC e agora membro do Comitê Executivo do ICSU, que coordenou o Grupo de Trabalho Open Access, inaugurou a coluna “ICSU News” do boletim bimestral IMU-Net.

A GDML

A visão que se tem da GDML é a de um corpo crescente de informação matemática, serviços web e agentes de software³, de domínio público e licença aberta, coexistindo com as publicações matemáticas e serviços de indexação atuais. Imagine *ser capaz de procurar na literatura por situações onde algum conceito específico de matemática foi usado ou resolvido*: isto permitiria considerar abordagens alternativas para solucionar suas próprias questões de pesquisa. Esse tipo de capacidade de busca poderia ser facilitado pelo uso de uma base de dados de informação gerada por máquinas e cultivada por seres humanos, sobre a literatura matemática, e permitiria a construção de uma série de outras capacidades.

Hoje temos a oportunidade de expandir e redefinir a forma como o conhecimento matemático é representado e usado, o caráter da literatura matemática e como ele evolui, e dar poder aos matemáticos pelas novas possibilidades de interação com o conhecimento. Esse futuro relacionamento com a literatura e com o corpo de conhecimento matemático irá além de novas formas de acesso e de ferramentas analíticas; ele também incluirá ferramentas e serviços para acomodar a criação, o compartilhamento e a pesquisa de novos tipos de estruturas de conhecimento.

³ N.E.: O termo original, ‘software agent’ é usado em ciência da computação para designar um programa de computador com objetivo definido que reage ao ambiente e roda sem supervisão direta e contínua, para realizar alguma tarefa para um usuário final ou para um outro programa. É um análogo de um robô autônomo. Ele pode se ativar e rodar por si próprio, sem que haja interação com um usuário humano, ou iniciar, supervisionar e fechar outros programas ou agentes.

Quais dados matemáticos?

Para o pesquisador em matemática, e para o cientista em geral, um dado importante é o conhecimento prévio. Esforços de digitalização do conhecimento matemático foram feitos nas últimas décadas com a European Digital Mathematics Library (EuDML), e a National Science Digital Library (NSDL), sendo estes exemplos proeminentes de serviços centralizados apoiados por metadados⁴. Atualmente, o conjunto de recursos digitais de informação bibliográfica para a literatura matemática é extenso:

- o MathSciNet (1940-hoje) tem aproximadamente 3 milhões de itens e indexa em torno de 2000 periódicos;
- o Zentralblatt, zbMATH (1868-hoje, incl. Jahrbuch), contém mais de 3 milhões de entradas e atualmente indexa mais de 3000 periódicos e séries.

Olhando para o crescimento de artigos de matemática adicionados nos últimos 5 anos ao arXiv, parece que cada vez mais a literatura matemática estará em forma digital, muitas vezes com formatação de alta qualidade, especificamente quando “nascida” digital ou retrodigitalizada para estar em um formato legível por máquinas, como o \LaTeX e o MathML.

Listas e tabelas sempre foram essenciais para o matemático. As mais básicas são as tabelas numéricas (por exemplo, valores de logaritmos, funções trigonométricas, funções especiais, zeros da função zeta, sequências de inteiros). Mais sofisticadas são listas de objetos matemáticos (por exemplo, integrais definidas e indefinidas, grupos simples finitos, transformadas de Fourier, equações diferenciais parciais e suas soluções), ou mesmo listas de definições, axiomas e teoremas. Coleções digitais desses tipos incluem:

- LOCOMAT (Loria Collection of Mathematical Tables): uma biblioteca de tabelas históricas recalculadas;
- NIST Digital Library of Mathematical Functions (DLMF): uma revisão digital do Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, de Abramowitz e Stegun, provavelmente a publicação técnica do National Institute of Standards and Technology⁵ mais amplamente distribuída de todos

⁴ N.E.: Metadados são dados sobre outros dados.

⁵ N.E.: Nome que passou a ter, em 1963, o National Bureau of Standards, fundado em 1901 nos EUA.

os tempos;

- On-Line Encyclopedia of Integer Sequences (OEIS): celebrando 50 anos de identificação de sequências e ainda recebendo em torno de 100 novas sequências e atualizações por dia, todas feitas por editores voluntários;

- Wolfram Functions Site: a coleção mais enciclopédica do mundo de informação sobre funções matemáticas.

Quais serviços matemáticos?

Modelagem e simulação têm sido aplicadas em muitas áreas da ciência, de aerodinâmica a economia e imagens médicas. Programas importantes para o cientista computacional baseiam-se em programas matemáticos como:

- Portable, Extensible Toolkit for Scientific Computation (PETSC): usado extensivamente em modelagem e simulação;

- SWMath: banco de dados extenso de informações sobre softwares matemáticos;

- TIDES: permite integrar problemas de EDO numericamente usando séries de Taylor com precisão múltipla, permitindo obter soluções com qualquer precisão;

- SuperLU: biblioteca de alto nível para a solução direta, em máquinas de alta performance, de sistemas de equações lineares grandes, esparsos e não simétricos.

Simulações para a modelagem climática e para o movimento planetário, ambos para grandes escalas e tempo longo, são exemplos de situações que podem requerer aritmética de alta precisão, além da precisão de 64 bits. Pacotes de software desenvolvidos recentemente atacaram esse desafio de computação científica fazendo da precisão numérica uma componente obrigatória tão importante para a programação quanto são os algoritmos e estruturas de dados.

Em conexão com a confiabilidade da ciência, a reprodutibilidade de resultados científicos tem recebido atenção e requer lidar com o ambiente de trabalho inteiro do cientista que faz os cálculos:

- Recomputation.org: um primeiro repositório para experimentos em ciência da computação atinge repro-

dutibilidade de resultados por “recomputação” e arquivamento de máquinas virtuais junto com código e dados.

Um grande passo à frente na descoberta de serviços matemáticos resultará também de melhorias na documentação de softwares matemáticos, que ainda são basicamente voltados para seres humanos e amplamente não categorizados ou não contendo qualquer tipo de metadados legíveis por máquinas.

Perspectivas futuras

A expansão do conjunto de programas de computador que usam matemática permitirá passar da mineração de texto para a mineração de matemática. Esse trabalho envolverá aspectos de classificação e representação do conhecimento matemático, linguística computacional, agregação e análise de *corpora* textuais, ferramentas para (meta)dados e processamento de textos integrais, OCR⁶ e análise de documentos, desenvolvimentos de recuperação de informações, e *workflows* de processamento de documentos. As necessidades de grandes comunidades de usuários guiarão esse processo na direção das metas da GDML e isto acabará por aumentar o acesso para todos.

Referências

- [1] National Research Council, Developing a 21st Century Global Library for Mathematics Research. Washington, D. C., 2014. ISSN: 978-0-309-29848-3
- [2] The World Digital Mathematics Library (WDML). blog.wias-berlin.de/imu-icmpanel-wdml

⁶ N.E.: OCR é o Reconhecimento Óptico de Caracteres.