

Sistema de Apoio e Gestão de Competências aplicado na Plataforma Lattes

Lucas dos Santos Abreu¹, Luiz Melo Romão¹

¹Departamento de Informática – Universidade da Região de Joinville (Univille)
Caixa Postal 246 - CEP 89219-710– Joinville – SC – Brasil

{lucas.abreu, luiz.melo}@univille.br

Resumo. *O conhecimento sobre a produção científica de uma instituição de ensino superior possui um papel importante para determinar a qualidade da mesma, assim como responder a oportunidades dadas a mesma seja por editais de licitação ou projetos de pesquisa. Com essas informações é possível determinar os limites e competências da instituição. Uma forma de analisar essas características é através dos currículos de seus profissionais. Nesse artigo serão indicados os conhecimentos necessários e métodos possíveis para alcançar estes indicadores de forma automatizada e orientada a tomada de decisão usando os dados dos currículos científicos da plataforma Lattes do CNPq.*

Abstract. *The knowledge about a High School Institution's scientific production it is important to define institution's quality itself, and to act on the opportunities that it will attempt, being biddings or research projects. With that information, it is possible to define the limitations and skills of the institution. One way to analyze those characteristics is using its professional's curriculums. In this article will be indicated the needed knowledge and possible methods to reach these indicator in an automatic and decision taking way using the data from the scientific curriculum from CNPq's Lattes platform.*

1 Introdução

As Instituições de Ensino Superior (IES) são avaliadas, entre outros requisitos, pela quantidade e qualidade de suas produções científicas. Dessa forma é importante para a IES manter atualizada a sua base sobre quantos e quais foram as produções geradas pelos seus professores, detalhando as linhas de pesquisa, áreas e qualidade das mesmas. Além de mensurar a qualidade da instituição é possível usar tais dados para responder a projetos e licitações que são oferecidas através da assessoria de prestação de serviços, definindo se é possível atender a essa oportunidade e quem são os profissionais qualificados para elas. Internamente esse conhecimento se torna importante para compreender quais projetos estão sendo executados na universidade e se não é possível aproximar tais projetos para alcançar resultados ainda maiores.

Estas atividades são muito importantes para a instituição e a maior parte dos dados usados para executá-las vem da base de currículos Lattes. Entretanto, muitas vezes esta coleta é feita de forma manual, o que, devido ao volume dos dados, leva bastante tempo, possibilitando que os dados se tornem defasados e obrigando a consultas constantes a base original – currículo a currículo.

Isso pode levar a perda de oportunidades para a instituição e no eventual desperdício do tempo de um profissional que poderia estar realizando outras atividades. Existem algumas ferramentas disponíveis que facilitam este processo como (Stela, 2015) e (Mena-Chalco, 2009). Porém, estas ferramentas possuem um custo elevado ou então não são tão simples de se utilizar.

Com esta pesquisa pretende-se mostrar como construir uma ferramenta que automatize a extração dos dados destes currículos e os disponibilizar em uma base de dados consolidada para que dessa forma possam estar disponíveis de forma simples, rápida e confiável para apresentação de indicadores e tomada de decisão.

Levando em conta essa situação, este trabalho propõe o desenvolvimento de um sistema de apoio e gestão de competências aplicado na Plataforma Lattes, a fim de automatizar o levantamento e apresentação das informações dos pesquisadores contidos na Plataforma Lattes de forma rápida e simples. Para construir tal ferramenta foram utilizados conceitos de Descoberta de Conhecimento (KDD), que permite criar um processo que busca, transforma e carrega os dados para uma base consolidada que fornece uma rápida consulta e interpretação dos dados existentes nos currículos. Como os dados da produção científica é relacionado às competências dos pesquisadores, também foram empregados conceitos provenientes da Gestão do Conhecimento (KM), dessa forma direcionando a estrutura da base de dados para facilitar a busca das competências dos profissionais e facilitar atividades relacionadas à seleção dos mesmos para projetos de pesquisa. Uma vez que a base de dados precisa ser concisa, de rápido retorno, e sem onerar outros sistemas, também foram empregados conhecimentos sobre *Data Warehousing* e *Business Intelligence* para que a leitura e pesquisa da base fosse a mais produtiva possível e direcionada para a tomada de decisão.

Além desta seção introdutória o artigo esta separado da seguinte forma: na seção 2 serão introduzidos conceitos gerais sobre este trabalho, na seção 3 estará descrita a metodologia utilizada neste trabalho, na seção 4 demonstraremos os resultados obtidos com este trabalho e na seção 5 concluímos a análise do trabalho e propomos futuras ações sobre o mesmo.

2 Referencial Teórico

O KDD (*Knowledge Discovery in Databases*) ou Descoberta de Conhecimento surgiu da união de áreas relacionadas ao tratamento de grandes massas de dados, buscando nelas padrões validos e novos com potencial para gerar dados compreensíveis e uteis (Fayyad, 1996).

Para alcançar seu objetivo o KDD se tornou um processo composto de etapas iterativas numa sequencia específica, mas que podem retornar a etapas anteriores a fim de realizar revisões ou ajustes no processo (Han, 2011). Essas etapas são oito e podem ser observadas na Figura 1:

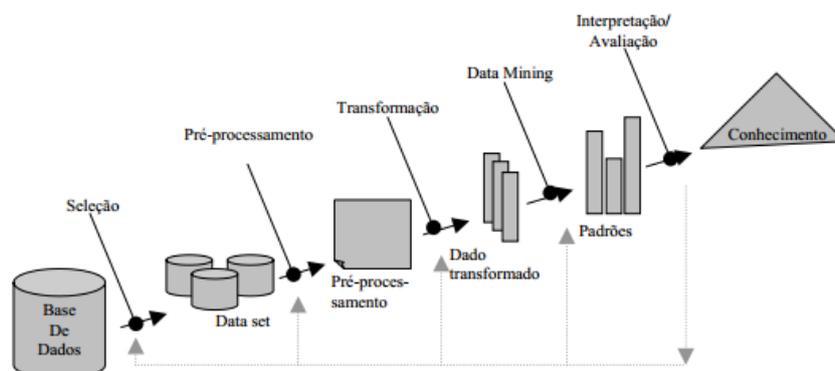


Figura 1 – Etapas que compõe o processo de KDD (PIATETSKY-SHAPIRO, 1996).

Estas etapas podem ser definidas da seguinte forma segundo Fayyad (1996):

1. Seleção: desenvolver uma compreensão do domínio (objetivo ou assunto dos dados), conhecimento prévio do conhecimento, e assim identificar o objetivo do KDD;
2. Data Set: Selecionar os dados onde o conhecimento provavelmente será encontrado;
3. Pré-processamento: Realizar um pré-processamento dos dados para tratar ruídos nos dados e lidar com eventuais “faltas” de dados no modelo.
4. Transformação: nessa etapa é feita agrupação dos dados ou transformação dos mesmos para reduzir o número de variáveis a considerar nos dados. Gerando os dados transformados.
5. Data Mining: Identificar o método de data mining que será utilizado (sumarização, classificação, regressão, etc.). Escolher um algoritmo de data mining para procurar os padrões nos dados
6. Executar a procura dos padrões de interesse dentro dos dados com base nas informações e modelos definidos previamente.
7. Interpretação/Avaliação: essa etapa envolve visualizar os dados, padrões e modelos para verificar a efetividade da solução, com isso pode ser necessário voltar para alguma etapa anterior para correção ou melhoria do modelo construído.
8. Conhecimento: consolidar os dados, enviando os mesmos para outros sistemas para uso, ou documenta-los e reportar para as partes interessadas no mesmo. Dessa forma tomando uma atitude com os dados extraídos ou lidar com as diferenças encontradas entre o esperado e obtido do modelo.

De acordo com Wiig (2000) o *Knowledge Management* (KM) ou Gestão do Conhecimento é uma área de conhecimento relaciona a gestão sistêmica das atividades, práticas e políticas de uma empresa, focando no conhecimento gerado e usado. Esse gerenciamento é feito se usando de ferramentas que retenham, analisem, organizem, melhorem e compartilhem o conhecimento de negócio existente nas atividades da empresa (Groof, 2012).

Embora as técnicas de KM sejam normalmente adicionadas às atividades cotidianas da empresa, uma boa gestão fará com que mesmo o aumento das atividades se transforme em menos trabalho e decisões mais rápidas e precisas (Wiig, 2000).

Wiig (2000) afirma que, com o uso das técnicas de KM, incentiva-se a colaboração dos profissionais, mantendo conhecimento dentro da empresa e melhorando a autoestima, uma vez que os mesmos passam a ser reconhecidos como fonte de valor em conhecimento.

Com a grande quantidade de informações disponíveis, o uso de ferramentas para armazenar e interpretar estas informações são necessárias. Conforme Negash (2004) *Business Intelligence* é uma ferramenta que permite transformar informação útil proveniente de diversos sistemas em conhecimento através da análise humana, essa análise pode ser feita por previsão e estimativa, análise de impacto baseadas em cenários, acesso rápido a dados para responder perguntas não rotineiras e insights estratégicos.

Para ser possível realizar tais análises com dados de vários sistemas é necessária à construção de um banco de dados unificado com as mesmas, a esse banco dá-se o nome *Data Warehouse* (DW).

As características que se destacam para esse tipo de banco de dados são: provem visão total e integrada da empresa; torna simples o uso de dados históricos e corretos na tomada de decisão; permite executar consultas pesadas sem onerar sistemas transacionais; mantem as informações organizadas e consistentes; e é uma fonte de informação estratégica flexível e interativa (Ponnian, 2004).

Já existem outros programas que tentam trazer resultados semelhantes aos alcançados com a solução proposta. Sendo que dois deles possuem maior destaque.

Um deles é o ScriptLattes que foi desenvolvido por Jesús P. Mena-Chalco e Roberto M. Cesar-Jr para auxiliar na avaliação dos currículos da Secretaria de Pós-graduação do IME-USP, o mesmo tem seu código fonte aberto e entrega vários relatórios estáticos sobre a produção técnica e bibliográfica dos professores e analisa relacionamentos de orientação e coautoria que existam nos currículos Lattes importados pela mesma (Mena-Chalco, 2016).

Outro programa é o Stela Experta que auxilia no apoio estratégico nas áreas de gestão do ensino, pesquisa, extensão e inovação e foi desenvolvido pelo instituto Stela, a mesma se utiliza das informações existentes na Plataforma Currículo Lattes para gerar indicadores relacionados à localização de expertises, redes de relacionamentos, mapas de conhecimento agrupando os dados em três temas Pessoas, Produções e Projetos e entregando gráficos e relatórios dentro desses temas (Stela, 2015).

3 Procedimentos Metodológicos

Nesta pesquisa foram analisados os dados qualitativos gerados pela inclusão de uma ferramenta de BI, uma vez que existem poucos trabalhos no mesmo tema, a execução teve um formato exploratório e de estudo de caso sobre a abordagem usada para extração e consulta dos dados para as áreas que fazem uso dos mesmos.

Para acessar os dados dos currículos, utilizou-se o acesso disponibilizado pelo CNPq via VPN que através de um arquivo em formato XML, permite a consulta das informações dos currículos Lattes dos professores da instituição.

Com base na documentação disponibilizada pelo CNPq, mapeou-se os dados presentes no XML e modelou-se uma base *Data Warehouse* com os dados dos currículos, porém mais acessível para consulta. Por exemplo, todas as Produções

Técnicas foram agrupadas em uma tabela que possui um identificador detalhando se esta produção é uma “Cultivar”, “Patente”, “Software”, etc., o que facilita a leitura dos dados em comparação a estrutura hierárquica do XML. Este tipo de “remodelagem” foi aplicado a praticamente todas as tabelas de modo que a base se tornasse mais performática e legível.

Outra informação importante incluída no *Data Warehouse*, foi das classificações dos periódicos do Qualis, criando a possibilidade de análise dos artigos gerados pela Instituição.

Nesta modelagem optamos por montar uma estrutura *constellation schema* para o *Data Warehouse*, que se caracteriza por existirem várias tabelas fato que se interligam pelas dimensões. Optou-se por este formato, pois existem fatos muito distintos e não necessariamente interligados nos currículos como, por exemplo, Bancas de Trabalho, Produção Bibliográfica e Participação Em Eventos, professores que participaram de bancas podem nunca ter feito uma produção bibliográfica e vice versa, utilizar uma tabela fato agregando tanto conteúdo poderia acarretar em um alto nível de complexidade para os usuários identificarem onde esta a informação que precisam.

Com a base modelada e acesso ao serviço do Lattes, foi desenvolvido um programa em C# para se comunicar com o serviço citado. A aplicação busca o currículo dos pesquisadores da IES e transforma os dados do XML para que possam ser gravados no *Data Warehouse*.

A aplicação ainda avalia as informações das produções científicas e projetos dos professores para que não fiquem duplicadas na base de dados baseando-se no ano, tipo e título das mesmas. Dessa forma, caso dois pesquisadores tenham citado um mesmo projeto ou produção em seus currículos, ao invés de inserir dois registros o programa irá apenas vincular os dois pesquisadores ao um mesmo registro de produção/projeto. Para a produção bibliográfica também é associado qual o extrato Qualis o mesmo está classificado.

Outras questões também foram implementadas, como limite de caracteres e valores vazios (nulos) são tratadas e por fim o programa extrator carrega os dados dos currículos na base de consulta completando o processo de ETL (*Extraction Transformation and Load* – extração, transformação e carga de dados). Mais detalhes sobre as regras podem ser verificadas diretamente no fonte do extrator em <https://fabricadesoftwareuniville.github.io/ExtratorLattesCNPq/> que estão sendo disponibilizados sobre a licença GNU GPL.

Uma vez que os dados estejam carregados no DW, utilizou-se a ferramenta de *Business Intelligence* Tableau, que permite a criação de painéis utilizando a base de dados construída anteriormente. No Tableau foram então criadas consultas e painéis para análise das informações geradas.

Os indicadores e painéis foram escolhidos baseando-se na análise de outras ferramentas no mercado e dos indicadores que as mesmas se propõem a entregar. Conjuntamente foi feita a discussão com os usuários para refinar as análises a serem construídas.

A cada execução o extrator verifica se o currículo em questão sofreu alterações desde a última extração e irá carregar apenas os currículos que foram alterados, dessa forma torna a execução incremental e mais performática.

4 Análise de dados

Uma vez que o programa de extração foi concluído e a base de consulta alimentada é possível realizar análises sobre os dados dos currículos de forma simples. A seguir serão descritas algumas consultas geradas com a ferramenta Tableau.

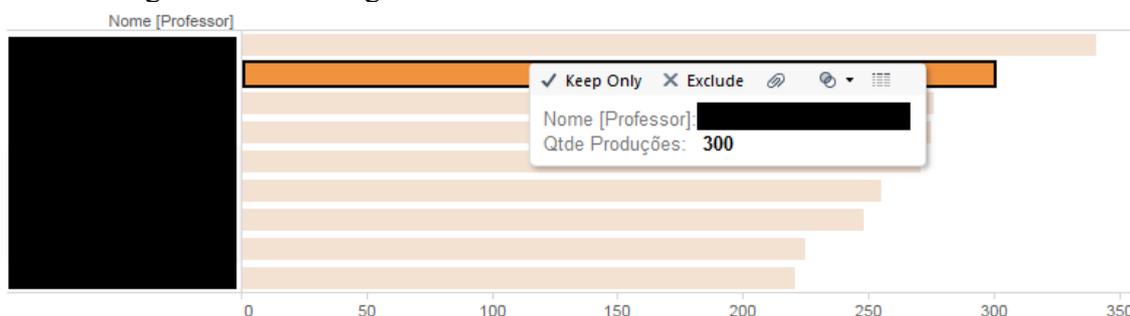


Figura 2 – Consulta Qtde Produções Bibliográficas X Professores

A Figura 2 mostra a quantidade de produções bibliográficas geradas pelos professores da instituição, esta informação por mais simples que seja tem grande valor para instituição uma vez que demonstra quanto conteúdo foi gerado para a comunidade científica e a produtividade de seus pesquisadores.

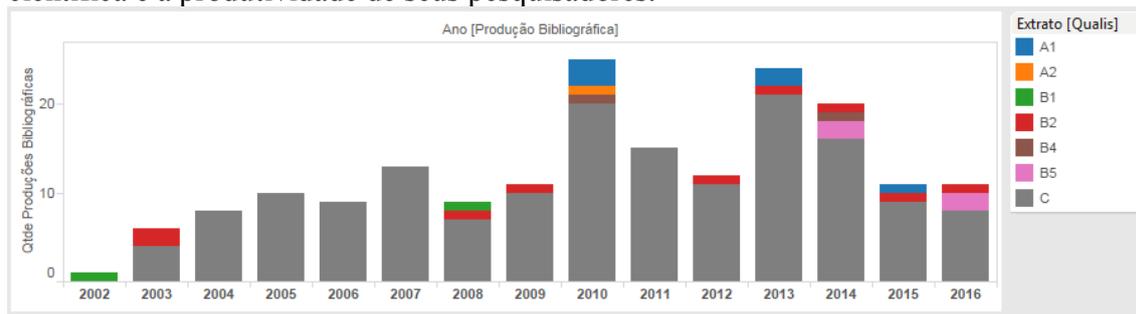


Figura 3 - Extrato Qualis na Produção Bibliográfica para Ciências da Computação

Na Figura 3 pode-se visualizar a distribuição da produção bibliográfica da instituição ao longo dos anos para a área de Ciências da Computação dentro do extrato Qualis, dessa forma permitindo constatar visualmente se a produção da instituição tem alcançado periódicos com maior ou menor notoriedade.

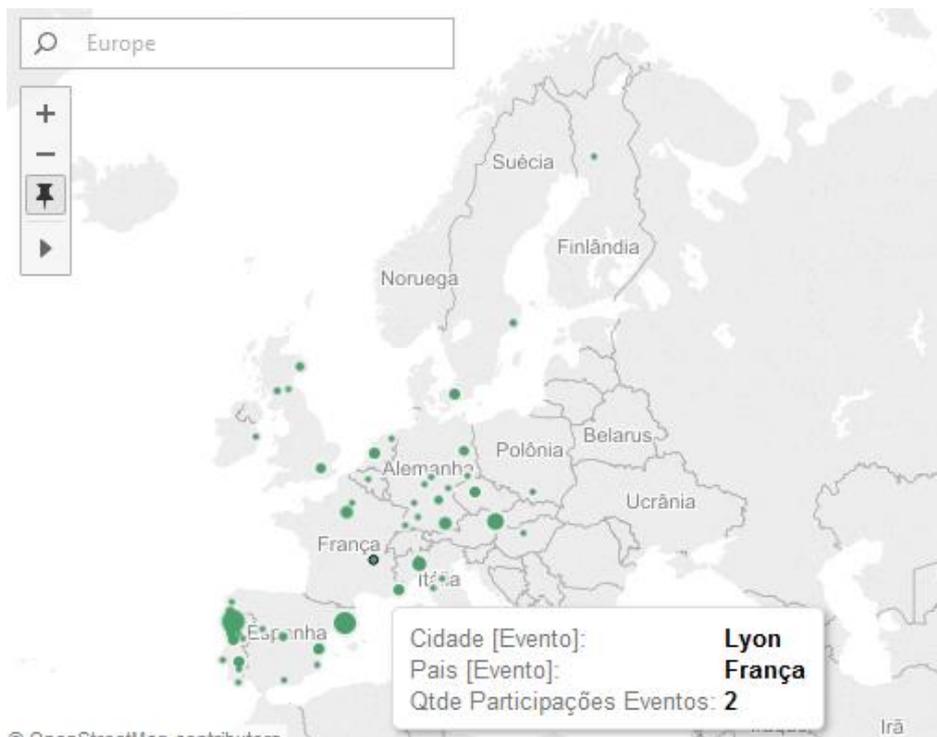


Figura 4 - Eventos Participados na Europa até 2015

Com a Figura 4 é possível visualizar a participação dos pesquisadores da IES na Europa até 2015, em nível de cidade. O tamanho do círculo também permite perceber a quantidade de eventos e pesquisadores que foram a uma determinada cidade. Também é possível ir sobre o ponto e ver a descrição específica do mesmo.

Também foram criados painéis com as consultas geradas, conforme mostra a figura 5.

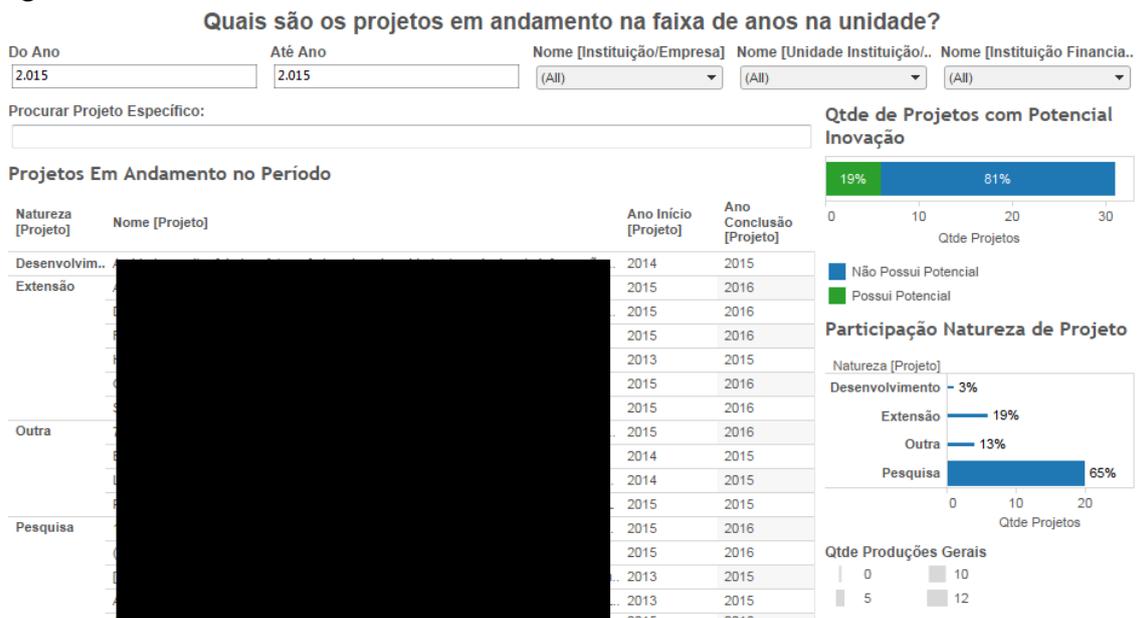


Figura 5 - Painel de Projetos

No painel demonstrado na Figura 5 estão compiladas várias informações sobre os projetos desenvolvidos pela IES. Uma listagem detalhando quais são os projetos em andamento no período pesquisado, qual a relação entre projeto com potencial de inovação e quantidade e percentual de projetos por categoria, sendo que todos estes podem ser filtrados pelos controles no topo do painel.

Embora os painéis destacados estejam com foco para a pesquisa a base construída nesse projeto também abrange atividades voltadas para a docência e administrativas dentro da comunidade científica, como Orientação, Bancas, Formação Acadêmica e Atividades Profissionais que estão informadas nos currículos dos profissionais.

5 Considerações Finais

Internamente todas as IES devem identificar de seus resultados quais são os pontos fracos e fortes das mesmas e então definir caminhos para superar seus problemas e refinar suas qualidades.

Um requisito necessário para avaliação das IES é a quantidade e qualidade da produção científica produzida, estes dados normalmente estão disponíveis na plataforma de currículos Lattes dos pesquisadores. Entretanto a coleta dos dados tende a ser manual em boa parte das IES e currículo a currículo, tornando a coleta morosa e suscetível a erros.

Neste trabalho foi proposto a construção de uma ferramenta para automatizar a busca e extração das informações dos currículos dos pesquisadores da instituição para alimentar uma base de consulta. Com a utilização de um software de BI é possível a análise dos dados e apoiar na gestão das atividades da organização.

Os resultados deste artigo já permitem um ganho no tempo e qualidade das informações coletadas. Análises que levavam quase três meses para serem concluídas podem ser entregues em minutos.

Para melhorar ainda mais os resultados obtidos pretende-se avançar com trabalhos para identificar de forma simples e automática os grupos de pesquisadores que existem organicamente na instituição, dessa forma contribuindo para definição de novas equipes de projetos; e avaliar o uso de outras ferramentas de BI que sejam gratuitas para aumentar o alcance do programa criado.

6 Referencias

CNPq. EXTRAÇÃO DE DADOS. Disponível em: <<http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados;jsessionid=E265541618ED4900C9AECCB88074061B>>. Acesso em: 12 jul. 2015

DE OLIVEIRA, Maria da Penha; CASSIS, Márcia Ramos. MAPEAMENTO DE COMPETÊNCIAS: Informações e Etapas Desenvolvidas para Construção Da Matriz-Base. Disponível em: <http://www.planejamento.gov.br/secretarias/upload/Arquivos/seges/EPPGG/outros/produto_V_mapeamento_competencias.pdf>. Acesso em: 24 mai. 2015.

FAYYAD, Usama M. et al. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: KDD. 1996. p. 82-88.

- GROFF, Todd; JONES, Thomas. Introduction to knowledge management. Routledge, 2012.
- GOEBEL, Michael; GRUENWALD, Le. A survey of data mining and knowledge discovery software tools. ACM SIGKDD Explorations Newsletter, v. 1, n. 1, p. 20-33, 1999.
- HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data mining: concepts and techniques: concepts and techniques. Elsevier, 2011.
- HAUSSLER, David; FAYYAD, Usama M.; STOLORZ, Paul E. KDD for Science Data Analysis: Issues and Examples. In: KDD. 1996. p. 50-56.
- MENA-CHALCO, Jesús P. e CESAR-JR, Roberto M. ScriptLattes: Uma ferramenta para extração e visualização de conhecimento a partir de Currículos Lattes. Disponível em: <<http://scriptlattes.sourceforge.net/description.html>>. Acesso em: 6 de jun. de 2016
- NEGASH, Solomon. Business intelligence. The Communications of the Association for Information Systems, v. 13, n. 1, p. 54, 2004.
- PIATETSKY-SHAPIRO, Gregory; FAYYAD, Usama; SMYTH, Padhraic. From data mining to knowledge discovery in databases. AI magazine, v. 17, n. 3, p. 37, 1996.
- PONNIAH, Paulraj. Data warehousing fundamentals: a comprehensive guide for IT professionals. John Wiley & Sons, 2004.
- SANTIAGO JR, José Renato Sátiro. Gestão do conhecimento. São Paulo: Novatec Editora, 2004.
- STELA. Plataforma Stela Experta .Disponível em: <<http://www.stelaexperta.com.br> > Acesso em: 29 fev 2015.
- WIIG, Karl M. Knowledge management: an emerging discipline rooted in a long history. Knowledge horizons: the present and the promise of knowledge management, p. 3-26, 2000.