

SUMÁRIO EXECUTIVO DE KNOW HOW

UNIVERSIDADE FEDERAL DE MINAS GERAIS-UFMG

A descoberta de informações não triviais, potencialmente úteis, significantes e reutilizáveis, implícitas em massas de dados, tem se tornado cada vez mais importante. O conjunto de técnicas computacionais capazes de realizar esta descoberta compõem o arcabouço ferramental da área de pesquisa e aplicação conhecida por mineração de dados (MD). Áreas como governo eletrônico, bioinformática, economia, sociologia, astrofísica, dentre outras, encontraram na MD muitas oportunidades de crescimento. Em especial, no cenário para governo eletrônico, onde encontramos grandes volumes de dados, é exigido das técnicas de MD o requisito essencial de escalabilidade.

A aplicação de técnicas de mineração de dados a bases de dados reais, como as descritas acima, demanda *algoritmos paralelos para mineração de grandes volumes de dados*. Geralmente esse processo é extremamente custoso do ponto de vista computacional, uma vez que a quantidade de dados envolvida na mineração em geral é muito grande, tende a crescer, e os algoritmos frequentemente tem complexidade exponencial. A utilização de soluções paralelas e distribuídas consolida-se como crucial para a viabilidade da execução desses algoritmos. Nesse sentido faz-se necessário uma *plataforma para paralelização de algoritmos*. Por outro lado, algoritmos de mineração de dados desenvolvidos em uma plataforma de paralelismo não são sustentáveis isoladamente. Logo, é preciso uma *plataforma integrada para serviços em mineração de dados* que proporcione todo um ambiente computacional para o processo de mineração de dados como serviços, tais como: controle de acesso, agendamento de tarefas, configuração de ambiente, gestão de hardware, dentre outros.

O *know-how* desenvolvido no âmbito do convênio intitulado “Práticas em Mineração de Dados Escalável para Sistemas de Governo” adquirido durante o período de 17/03/2011 a 16/09/2012 no contexto do convênio “Pesquisa Aplicada em Mineração de Dados para Governo Eletrônico” (RG nº 45.449), firmado entre o Serviço Federal de Processamento de dados (SERPRO) e a Universidade Federal de Minas Gerais (UFMG) através do Departamento de Ciência da Computação (DCC), em linhas gerais, é o conhecimento em *algoritmos paralelos para mineração de grandes volumes de dados*, na *plataforma para paralelização de algoritmos Anthill* e sob a perspectiva de uma *plataforma integrada para serviços em mineração de dados*. Ele engloba todo conhecimento adquirido na aplicação de técnicas de mineração de dados escalável em dados de sistemas governamentais.

DESCRIÇÃO DE KNOW HOW

UNIVERSIDADE FEDERAL DE MINAS GERAIS-UFMG

1 - Nome do Know-How

Este documento descreve o *know-how* intitulado “*Práticas em Mineração de Dados Escalável para Sistemas de Governo*” adquirido durante o período de 17/03/2011 a 16/09/2012 no contexto do convênio “Pesquisa Aplicada em Mineração de Dados para Governo Eletrônico” (RG nº 45.449), firmado entre o Serviço Federal de Processamento de dados (SERPRO) e a Universidade Federal de Minas Gerais (UFMG) com a participação de pesquisadores do Departamento de Ciência da Computação (DCC) vinculado ao Instituto de Ciências Exatas (ICEx).

2 - Descrição do Know-How

O *know-how* desenvolvido no âmbito do convênio SERPRO/UFMG, em linhas gerais, é o conjunto de *algoritmos paralelos para mineração de grandes volumes de dados*, a *plataforma para paralelização de algoritmos Anthill* e uma proposta de *plataforma integrada para serviços em mineração de dados*. Ele engloba todo conhecimento adquirido na aplicação de técnicas de mineração de dados escalável em dados de sistemas governamentais. Esse *know-how* pode ser expressado como a capacidade de reunir o conhecimento acadêmico em mineração de dados escalável e o conhecimento em governo eletrônico com o intuito de prover soluções para problemas relacionados a governo. A junção desses conhecimentos resulta em especialistas capazes de definir aplicações, modelos, algoritmos e sistemas.

No âmbito do convênio, o conhecimento foi produzido em três fases distintas: (1) pesquisa e desenvolvimento; (2) projetos piloto; (3) formação de disseminadores em mineração de dados escalável. Na fase de pesquisa e desenvolvimento foram exploradas técnicas de mineração de dados escalável em bases de dados de governo, mais especificamente dados de convênios derivados do Sistema de Transferências Voluntárias da União, SICONV (<http://www.convenios.gov.br/siconv>). Além disso, um novo algoritmo para agrupamento de texto foi proposto (MaxCluster). Já na segunda fase, o conhecimento adquirido inicialmente foi aplicado em três necessidades elicítadas tendo em vista a perspectiva de aplicação de mineração de dados no SICONV, resultando em três projetos pilotos. Por último, na terceira fase um processo de transferência do conhecimento adquirido foi realizada. Nessa etapa funcionários do SERPRO participaram de um curso de formação ministrado por docentes e discentes da universidade totalizando 80 horas ao longo de dois meses.

Ademais, é importante destacar a maturação do aprendizado em processos de descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases – KDD*) para governo eletrônico com ênfase na etapa de mineração de dados. Este conhecimento evoluiu de forma transversal aos conhecimentos produzidos nas fases mencionadas.

A Figura 1 ilustra um método iterativo, o qual descreve as práticas em mineração de dados aplicadas no âmbito do convênio SERPRO/UFMG. Estas práticas resultaram no *know-how* aqui explicitado.



Figura 1. Método iterativo para mineração de dados

Na primeira etapa da iteração, representada na Figura 1 pelo quadrante *aplicações*, demandas de mineração de dados são especificadas (normalmente na forma de perguntas de negócio). A partir dessas demandas, inicia-se a etapa de modelagem (quadrante *modelos*), onde soluções em nível conceitual para os problemas de mineração de dados são apresentadas. Naturalmente, os modelos escolhidos/desenvolvidos devem fundamentar a implementação de algoritmos (quadrante *algoritmos*). Por fim, estes algoritmos são refinados para execução, obedecendo restrições computacionais esperadas (quadrante *sistemas*). Na medida em que novas iterações de todo o processo são realizadas, espera-se um aumento na relevância das aplicações, expressividade dos modelos, utilidade dos padrões gerados pelos algoritmos e finalmente, um aumento na escalabilidade e eficiência dos sistemas gerados¹.

A seguir, os três pontos centrais desse *know-how*, o arcabouço para paralelização de algoritmos Anthill, um conjunto de algoritmos para mineração de dados escaláveis e uma plataforma integrada para serviços em mineração de dados, são discutidos:

2.1. Arcabouço para paralelização de algoritmos Anthill

¹ Outras qualidades de software podem ser aqui consideradas, tais como interoperabilidade, reuso, usabilidade, etc.

Anthill é um arcabouço para aplicações distribuídas, inspirado no modelo de programação filtro-fluxo desenvolvido pelo DCC/UFMG (<http://www.speed.dcc.ufmg.br/anthill>). Nesse modelo, a abordagem *data-flow* é utilizada para decompor as aplicações em um conjunto de unidades de processamento denominadas filtros. Dados são transferidos entre esses filtros via canais de comunicação permitindo que fluxos de registros de tamanho fixo sejam entregues para cada filtro.

O processo de decomposição da aplicação conduz á possibilidade de paralelismo de tarefas, na medida em que a aplicação torna se uma coleção de filtros independentes conectados formando um grafo direcionado. Em tempo de execução, múltiplas instâncias (transparentes) de cada um desses filtros podem ser instanciadas nos diversos processadores que constituem a plataforma de execução. Nesse cenário, cada mensagem no fluxo de entrada de cada filtro pode não precisar ser entregue em todas essas instâncias, constituindo, assim, paralelismo de dados. A Figura 2 apresenta uma visão esquemática da arquitetura do arcabouço Anthill para desenvolvimento de algoritmos distribuídos.

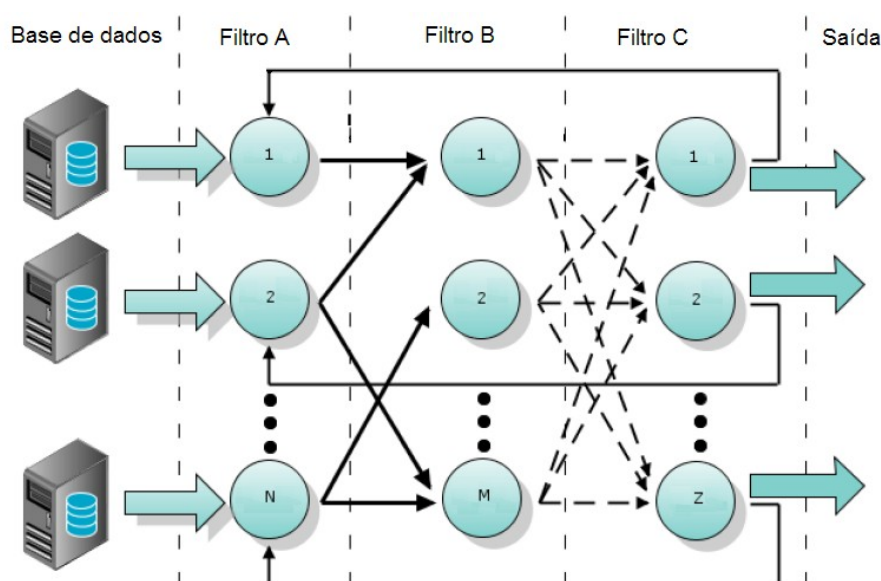


Figura 2. Visão esquemática da arquitetura do arcabouço Anthill.

Para muitas aplicações, a decomposição natural conduz a um grafo cíclico onde a execução corresponde a múltiplas iterações pelos filtros que constituem o ciclo. Quando a aplicação inicia, os dados de entrada representam o conjunto inicial de soluções candidatas que a aplicação tenta encontrar. Cada uma dessas soluções candidatas precisa ser então avaliada pelos filtros e na medida que isso vai sendo realizado, novas soluções candidatas são construídas que, por sua vez, precisam ser avaliadas. A experiência no desenvolvimento de aplicações Anthill mostra que essa abordagem leva a uma execução

assíncrona do ciclo, ou seja, múltiplas soluções candidatas são testadas simultaneamente em diferentes instâncias dos filtros. Esse comportamento é similar ao comportamento de software *pipelining* e paralelismo de *loop*.

Dessa forma, Anthill permite explorar o máximo de paralelismo da aplicação via a extração simultânea do paralelismo nas múltiplas dimensões discutidas, a dizer: paralelismo de tarefas, paralelismo de dados e assincronia. O resultado final é um paralelismo de grão bastante fino e de tarefas majoritariamente independentes, o que leva a alta escalabilidade.

2.2. Algoritmos para Mineração de Dados Escaláveis

O processo de mineração de dados demanda um conjunto de algoritmos que proporcione flexibilidade na execução de uma tarefa. No âmbito do convênio trabalhou-se com os seguintes modelos e algoritmos:

- ✧ Modelos para aprendizado de regras de associação
 - Algoritmos:
 - Apriori
 - Eclat - *Equivalence Class Transformation*
 - FP-growth.
- ✧ Modelos de agrupamento
 - Algoritmo:
 - MaxCluster
- ✧ Modelos de classificação
 - Algoritmo:
 - LAC – *Lazy Associative Classification*
- ✧ Modelos de mineração de grafos
 - Algoritmo:
 - SCPM – *Structural Correlation Pattern Mining*

Em especial, foi desenvolvido durante o convênio o algoritmo MaxCluster para realização de agrupamento textual, que corresponde a identificar grupos de registros textuais similares entre si. Cada registro é similar aos demais pertencentes ao mesmo grupo, e diferente dos registros pertencentes a outros grupos. O objetivo do agrupamento é encontrar registros similares entre si não conhecidos inicialmente. Esta tarefa é considerada desafiadora, sendo que os algoritmos existentes necessitam de uma grande

quantidade de parâmetros, fragmentam grupos indevidamente, ou simplesmente não escalam para grandes quantidades de dados. Na tentativa de dirimir tais limitações foi proposto o algoritmo MaxCluster.

O algoritmo MaxCluster emprega o *coeficiente de Jaccard* para determinar a similaridade entre todos dos registros. Inicialmente, gera-se uma matriz de similaridade entre todos os registros que deseja-se agrupar. Na sequência os grupos candidatos são enumerados pelo algoritmo. Nesta etapa, grupos contendo registros similares são fundidos e forma grupos cada vez maiores. Para limitar o espaço de busca apenas conjuntos maximais são considerados, i.e., um grupo não pode ser subconjunto de outro.

Para prover desempenho e escalabilidade o algoritmo MaxCluster foi paralelizado na plataforma Anthill. A estratégia de paralelização adotada utiliza o conceito de classes de equivalência. Mais especificamente, o espaço de busca por grupos maximais é particionado, e cada partição é atribuída a uma unidade processadora diferente. Dessa forma, a enumeração dos grupos maximais é realizada de maneira independente, sem a necessidade de sincronização entre os diversas unidades processadoras. Ao final da enumeração, os grupos que não forem maximais são removidos.

2.3. Plataforma Integrada para Serviços em Mineração de Dados

Na terceira fase do convênio um processo de formação de disseminadores em mineração de dados escalável foi realizado. Totalizando 40 horas ao longo de dois meses (06/2012 a 07/2012) os alunos se organizaram em grupos e foram instruídos a selecionar bases de dados reais mantidas pelo SERPRO, elaborar perguntas intensivas em conhecimento e respondê-las utilizando técnicas de mineração de dados. Durante essa fase, observou-se que algoritmos de mineração de dados desenvolvidos em uma plataforma de paralelismo não são sustentáveis isoladamente. Logo, é preciso uma *plataforma integrada para serviços em mineração de dados* que proporcione todo um ambiente computacional para o processo de mineração de dados como serviços, tais como: controle de acesso, agendamento de tarefas, configuração de ambiente, gestão de hardware, dentre outros. Além disso, o processo de mineração de dados naturalmente demanda um processo de experimentação; na verdade, MD é um processo iterativo composto de aplicações, modelos, algoritmos e sistemas. Uma vez consolidada uma análise, pode-se, caso necessário, migrar toda a configuração realizada para um ambiente de produção. A Figura 3 esboça um diagrama inicial desse ambiente. Inicialmente tem-se uma centralização de amostras de dados. De posse dessas amostras, cada projeto de MD instância uma *View*, isto é, um conjunto de dados para realização do processo de MD. Após obtenção de um modelo adequado para aplicação em mente o mesmo pode ser migrado para uma ambiente produtivo. Ainda na Figura 3 observe que diversos atores interagem nesse ambiente; cada um responsável por uma etapa do processo de mineração de dados.

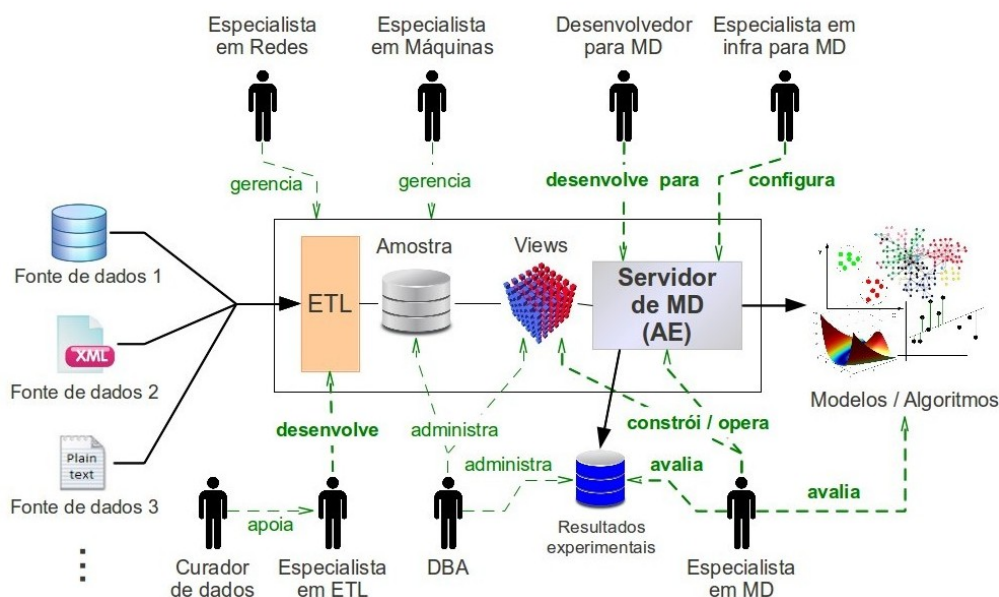


Figura 3. Diagrama esquemático de um ambiente produtivo em mineração de dados.

A Figura 4 concretiza o diagrama esquemática do ambiente produtivo em mineração de dados no diagrama de componentes da plataforma integrada para serviços em mineração de dados. A plataforma proposta é centrada em três componentes essenciais: algoritmos para mineração de dados e pré-processamento, *Pentaho Data Integration* e arcabouço para paralelização de algoritmos Anthill.

Pentaho data integration (PDI) (<http://kettle.pentaho.com/>) é um componente da suíte do Pentaho responsável pelos processos de extração, transformação e carga (ETL). Nele, os processos são criados em uma ferramenta gráfica que permite especificar o processo ETL sem a necessidade de escrever nenhum código. Como ferramenta de ETL, o PDI é a mais popular ferramenta *open source* disponível. Ele suporta um vasto conjunto de formatos de entrada e saída de dados, incluindo arquivos texto, xls e bancos de dados abertos e comerciais.

Associado ao componente PDI encontram-se componentes responsáveis pelo gerenciamento de algoritmos para mineração de dados escaláveis e pré-processamento. Já como suporte à execução dos algoritmos tem-se o arcabouço Anthill. Com o intuito de fornecer uma plataforma integrada para serviços em mineração de dados outros componentes importantes em um ambiente produtivo são apresentados; por

exemplo, controles de acesso, controles de visão, gestão de hardware, agendamento de tarefas, dentre outros.

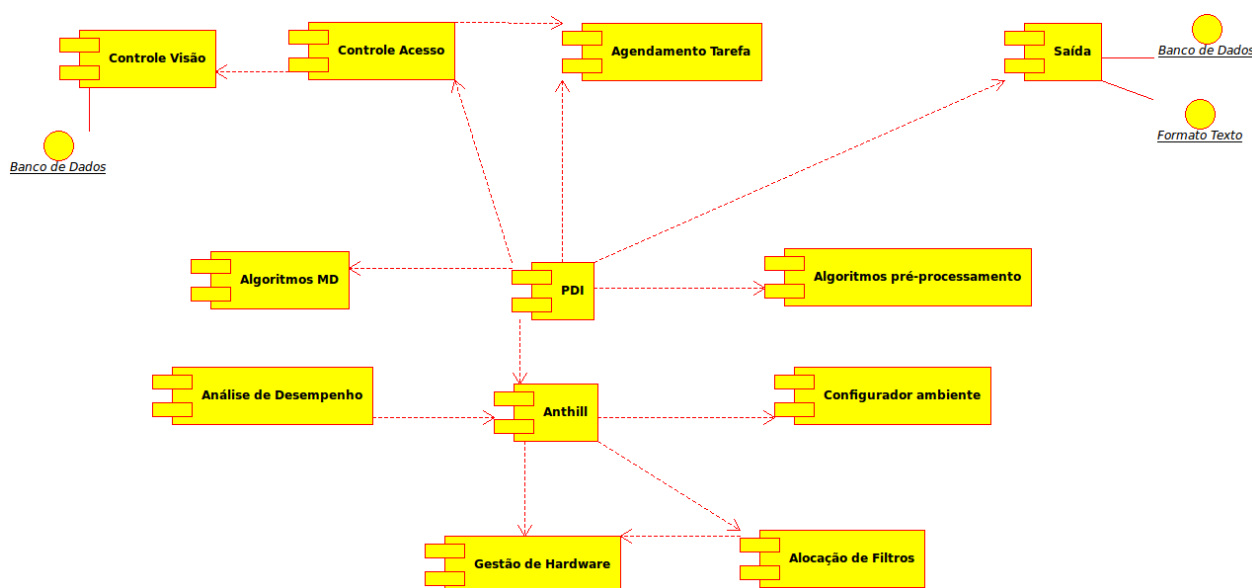


Figura 4. Diagrama de componentes da plataforma integrada para serviços em mineração de dados

3 - Inovação proposta e suas vantagens

O *know-how* descrito caracteriza um compartilhamento de conhecimento entre a UFMG e SERPRO no contexto de aplicação de técnicas de mineração de dados escalável em dados reais de sistemas de Governo (e.g. SICONV) implementada de acordo com processo iterativo de mineração de dados escalável (ilustrado na Figura 1).

Diante do corpo de conhecimento produzido ao longo do período que compreende entre 17/03/2011 e 16/09/2012 do convênio SERPRO-UFMG RG nº 45.449, concluímos que o mesmo pode representar vantagem estratégica para ambas as instituições. De um lado, o SERPRO enquanto prestador de serviços de TIC para o Governo Federal (e, em última análise, para o cidadão) pode se utilizar deste conhecimento para o enriquecimento dos seus serviços. Além disso, capacitar em novas linhas de negócio importantes para concretização e consolidação da transparência pública. De outro lado, a UFMG ganha conhecimento em como aplicar pesquisas desenvolvidas em mineração de dados e processamento paralelo de alto desempenho em cenários corporativos/industriais reais e de grande porte como os do SERPRO. Os dois

lados, analisados em conjunto, representam prática virtuosa sustentável para evolução do e-Gov no Brasil nesta dimensão.

4 - Aplicações e/ou usos

Essa seção descreve a aplicação do *know-how* supracitada em dados provenientes do SICONV (<http://www.convenios.gov.br/siconv>). O SICONV é um sistema desenvolvido em plataforma web que permite aos órgãos concedente e conveniente o gerenciamento *on-line* de todos os convênios cadastrados. Entende-se por convênio um acordo, ajuste ou qualquer outro instrumento que discipline a transferência de recursos financeiros de dotações consignadas nos Orçamentos Fiscal e da Seguridade Social da União e tenha como partícipe, de um lado, órgão ou entidade da administração pública federal, direta ou indireta, e, de outro lado, órgão ou entidade da administração pública estadual, distrital ou municipal, direta ou indireta, ou ainda, entidades privadas sem fins lucrativos, visando a execução de programa de governo, envolvendo a realização de projeto, atividade, serviço, aquisição de bens ou evento de interesse recíproco, em regime de mútua cooperação.

Relacionado a um convênio, os seguintes conceitos fazem-se necessários:

- ✦ **Contrato de repasse** - instrumento administrativo por meio do qual a transferência dos recursos financeiros se processa por intermédio de instituição ou agente financeiro público federal, atuando como mandatário da União;
- ✦ **Termo de parceria** - instrumento jurídico previsto na Lei 9.790, de 23 de março de 1999, para transferência de recursos para organizações sociais de interesse público;
- ✦ **Concedente** - órgão da administração pública federal direta ou indireta, responsável pela transferência dos recursos financeiros ou pela descentralização dos créditos orçamentários destinados à execução do objeto do convênio;
- ✦ **Contratante** - órgão ou entidade da administração pública direta e indireta da União que pactua a execução de programa, projeto, atividade ou evento, por intermédio de instituição financeira federal (mandatária) mediante a celebração de contrato de repasse;
- ✦ **Conveniente** - órgão ou entidade da administração pública direta e indireta, de qualquer esfera de governo, bem como entidade privada sem fins lucrativos, com o qual a administração federal pactua a execução de programa, projeto/atividade ou evento mediante a celebração de convênio;

- ✦ **Contratado** - órgão ou entidade da administração pública direta e indireta, de qualquer esfera de governo, bem como entidade privada sem fins lucrativos, com a qual a administração federal pactua a execução de contrato de repasse;
- ✦ **Proponente** - órgão ou entidade pública ou privada sem fins lucrativos credenciada que manifeste, por meio de proposta de trabalho, interesse em firmar instrumento regulado por esta Portaria.

Um dos problemas relatados por gestores públicos que precisam de informações gerenciais provenientes do Siconv é a falta de padronização nas descrições textuais (campos objeto e descrição) fornecidas pelos proponentes em tempo de proposta. O cenário ideal é aquele em que propostas de convênios que atacam objetivos similares possuam descrições textuais próximas, facilitando a tarefa de agrupamento de convênios, visando uniformizar tratamento para os mais variados fins. Frente essas necessidades e através do *know-how* adquirido na fase de pesquisa três projetos pilotos foram propostos e executados:

- ✦ *Análise de padrões de empenho em convênios* – descoberta de padrões de similaridade entre convênios com base em seus empenhos;
- ✦ *Geração de assinaturas de propostas de convênios* – geração de assinaturas de campos textuais de grupos de convênios;
- ✦ *Análise de redes sociais de concedentes e convenientes* – investigar os padrões de relacionamentos entre convenientes e concedentes.

Nas próximas subseções o processo empregado em cada projeto piloto é apresentado e discutido. Os dados utilizados nos experimentos são originados da base de dados do Siconv derivada das extrações sequenciais, contendo as tabelas “*convênios*” (convênios), “*empenhos*” (empenhos), “*programas*” (programas) e “*execfin*” (execução financeira).

Piloto 1 - Análise de Padrões de Empenhos em Dados de Convênios

Este piloto visa agregar valor às análises de dados de empenhos previstos para convênios, através da descoberta de padrões de similaridade entre convênios com base em seus empenhos. Cada convênio realiza um conjunto de empenhos ao longo de seu tempo de vida. Para convênios similares em termos de campos textuais, é esperado que eles possuam padrões similares de empenhos. O processo descrito a seguir visa detectar convênios que estão dentro e fora dos padrões de empenhos para um determinado grupo.

Agrupamento de Convênios por Similaridade Textual (Fase 1)

Em uma primeira fase, todos os convênios da amostra (registros cujo *campo Tx_Modalidade da tabela convenios tem valor igual a "Convênio"*) que já foram executados (registros cujo *campo Tx_Situação da tabela convenios tem valor que obedece ao padrão "*Prestação de Contas*"*) e que não têm empenhos diferentes dos do tipo "Empenho Original" (registros da tabela de convenios cujos *empenhos observados só possuam valor "Empenho Original" para o campo Tx_Tipo_Nota_Empenho na tabela empenhos*) são agrupados com base na similaridade entre o conjunto de termos extraídos dos seguintes campos:

- ✦ *Nm_Programa* (tabela programas);
- ✦ *Tx_Descricao_Programa* (programas);
- ✦ *Cd_Orgao_Superior* (programas);
- ✦ *Cd_Orgao_Concedente* (programas);
- ✦ *Cd_Executor* (programas);
- ✦ *Tx_Esfera_Adm_Proponente* (tabela convenios);
- ✦ *Tx_Regiao_Proponente* (convenios);
- ✦ *Uf_Proponente* (convenios);
- ✦ *Nm_Municipio_Proponente* (convenios);
- ✦ *Cd_Identif_Proponente* (convenios);
- ✦ *Tx_Objeto_Convenio* (convenio);
- ✦ *Tx_Justificativa* (convenios);
- ✦ *Cd_Respons_Proponente* (convenios);
- ✦ *Cd_Respons_Concedente* (convenios).

Os campos mencionados são usados para geração de um conjunto de termos para cada convênio. O processo de agrupamento é ilustrado na Figura 5. A figura esboça que um conjunto de registros em um espaço n-dimensional é segmentado formando grupos de convênios semanticamente semelhantes.

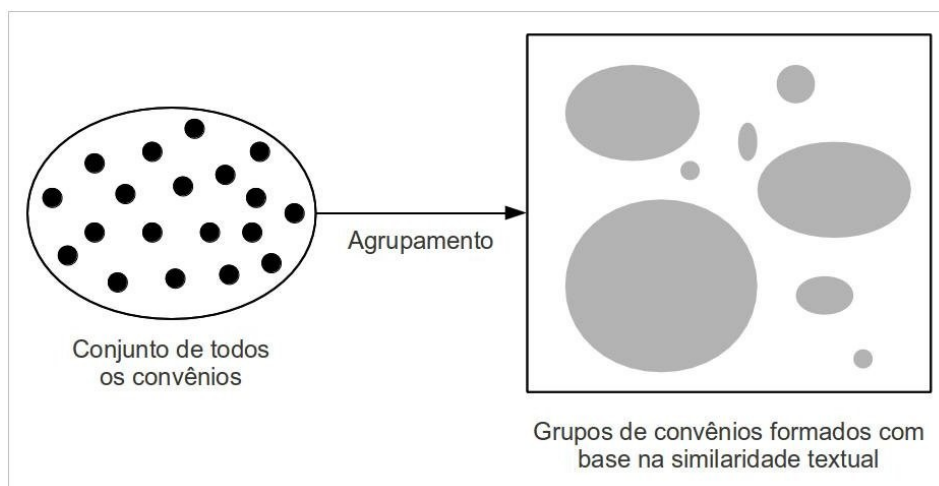


Figura 5. Agrupamento de convênios por similaridade textual.

Agrupamento de Convênios por Similaridade de Empenhos (Fase 2)

Em uma segunda fase convênios são agrupados pelos seus padrões de empenho. Cada um dos 764 possíveis valores de natureza de despesa (campo Cd_Natureza_Despesa da tabela empenhos) representa uma dimensão da base de dados. Em cada dimensão foram sumarizados os valores de todas as ocorrências de empenhos com a natureza de despesa correspondente à dimensão. Considerando o elevado número de dimensões, a base de dados resultante é uma matriz esparsa. O processo de agrupamento por similaridade de padrões de empenho é ilustrado na Figura 6, que esboça que um conjunto de registros em um espaço n-dimensional é segmentado formando grupos de convênios semelhantes em relação aos valores empenhados.

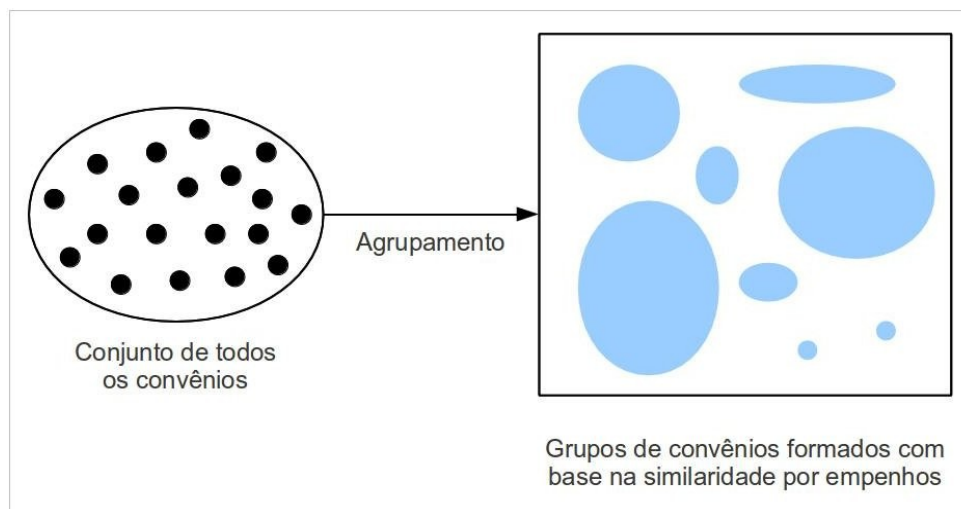


Figura 6. Agrupamento de convênios por similaridade de padrões de empenho.

Análise dos Grupos (Fase 3)

Em uma terceira e última fase, os seguintes questionamentos são colocados em evidência: (1) Há convênios similares textualmente, mas dissimilares em seus padrões de empenhos? (2) Há convênio dissimilares textualmente, mas similares em seus padrões de empenho?

O resultado para a primeira pergunta será uma lista de convênios cujos padrões de empenho estão fora do esperado. É importante notar que um convênio estar fora do padrão de empenho esperado pode indicar uma boa ou má gestão de recursos disponíveis, ficando a cargo de autoridades competentes a realização de auditorias e emissão de juízo de valor sobre as análises. O resultado para a segunda pergunta é uma lista de convênios que foram colocados em grupos não esperados com base nos padrões de empenho, valendo a mesma observação sobre juízo de valor sobre as análises. A Figura 7 ilustra a sobreposição de grupos para realizar este tipo de análise.

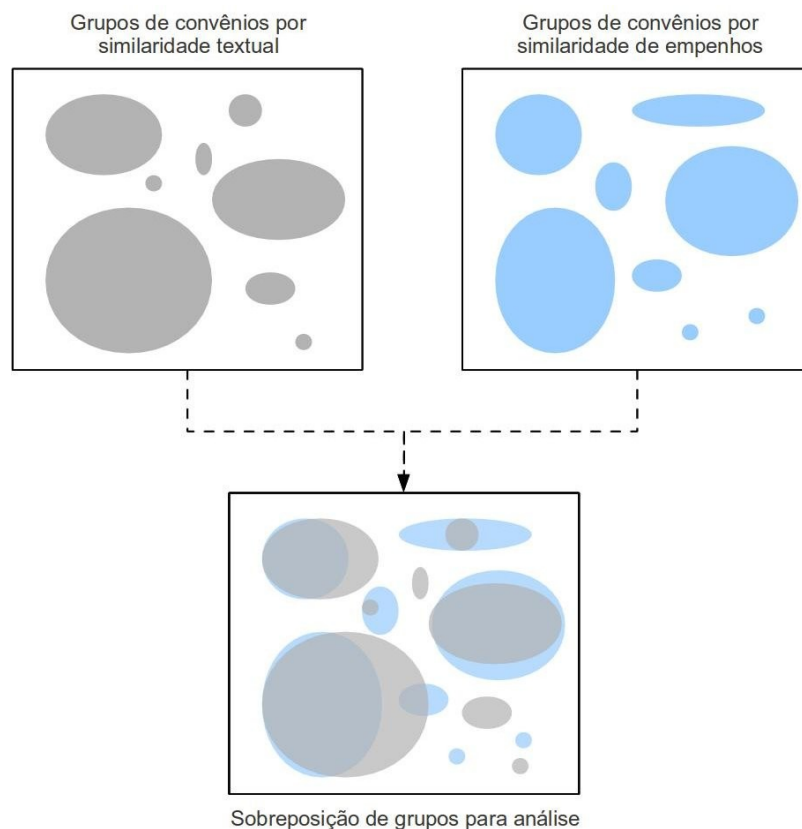


Figura 7. Sobreposição de grupos para análise de casamento entre padrões de descrição textual e padrões de empenho.

Experimentos

Agrupamento de Convênios por Similaridade Textual

Esta seção descreve as execuções ligadas ao agrupamento de convênios por similaridade textual, rotulada neste documento de Fase 1.

Preprocessamento

A seguinte consulta foi empregada para formação da base inicial da Fase 1:

```
select programas.Nm_Programa, programas.Tx_Descricao_Programa, programas.Cd_Org_Superior, programas.Cd_Org_Concedente,
programas.Cd_Executor, convenios.Tx_Esfera_Adm_Proponente, convenios.Tx_Regiao_Proponente, convenios.Uf_Proponente,
convenios.Nm_Municipio_Proponente, convenios.Cd_Identif_Proponente, convenios.Tx_Objeto_Convenio, convenios.Tx_Justificativa,
convenios.Cd_Respons_Proponente, convenios.Cd_Respons_Concedente from convenios, programas where
programas.Cd_Programa=convenios.Cd_Programa and convenios.Tx_Situacao like "%Prestação de Contas%" and
```

convenios.Tx_Modalidade="Convênio" and convenios.Nr_Convenio not in (select distinct(convenios.Nr_Convenio) from convenios, empenhos where empenhos.Nr_Convenio=convenios.Nr_Convenio and empenhos.Tx_Tipo_Nota_Empenho != "Empenho Original");

O resultado desta consulta é um conjunto de convênios que já foram executados e que não têm em seus históricos empenhos diferentes do tipo “Empenho Original”. Esta consulta representa a primeira fase do fluxo de trabalho definido do PDI exibido na Figura 8, gerando ao final um arquivo CSV com separador vírgulas, vírgulas removidas dos campos, espaços removidos no início e final dos campos, todos os caracteres em minúsculo, aspas obrigatórias em todos os campos, cabeçalho incluso, codificação UTF-8, formato de terminação de linha Unix LF e opção “*Fast data dump (no formatting)*” habilitada. Adicionalmente, caracteres especiais, por exemplo !, ?, * e &, também foram removidos.

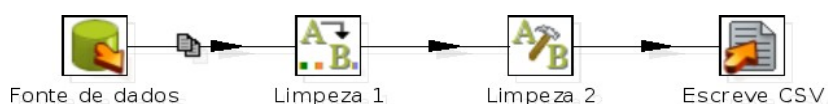


Figura 8. Pré-processamento inicial da Fase 1 realizado no PDI.

Para realização do processo de agrupamento optou-se pela utilização do algoritmo MaxCluster, desenvolvido no âmbito do convênio SERPRO-UFMG. Comparado com outros algoritmos para agrupamento de texto existentes na literatura, o algoritmo MaxCluster apresenta a vantagem de não necessitar de uma grande quantidade de parâmetros para sua calibração, não fragmentar grupos indevidamente e ser facilmente escalonável. Essa última característica é essencial ao lidar com grandes volumes de dados. A base de dados de entrada para o algoritmo de agrupamento MaxCluster é uma lista de conjuntos de termos, com identificador do conjunto no começo da linha (iniciando por 1). Esta base foi gerada removendo-se os caracteres de separação do arquivo CSV, as aspas que delimitam os *strings*, removendo-se algumas *stopwords* (Tabela 1) e colocando-se o identificador no começo da linha. O arquivo resultante é utilizado como entrada para o algoritmo MaxCluster. Um trecho deste arquivo é apresentado na Tabela 2.

Tabela 1. Lista de *stopwords* removidas durante o pré-processamento

a	esses	se
ainda	esta	seja
alem	estas	sem
ambas	este	sendo
ambos	estes	seu
antes	ha	seus
ao	isso	sob
aonde	isto	sobre
aos	logo	sua
apos	mais	suas
aquele	mas	tal
aqueles	mediante	tambem
as	menos	teu
assim	mesma	teus

com como contra contudo cujas cuja cujas cujo cujos da das de dela dele deles demais depois desde desta deste dispoe dispoem diversa diversas diversos do dos durante e ela elas ele eles em entao entre essa essas esse	mesmas mesmo mesmos na nas nao nas nem nesse neste no nos o os ou outra outras outro outros pelas pelas pelo pelos perante pois por porque portanto proprio proprios quais qual qualquer quando quanto que quem quer	toda todas todo todos tua tuas tudo um uma umas uns
--	---	---

Tabela 2. Exemplo de justificativa utilizada para agrupamento de convênios.

Id	Justificativa
209	fomento desenvolvimento turistico local chamada publica programa visa apoiar projetos incubacao empreendimentos economicos solidarios cadeia produtiva turismo projetos turismo base principios economia solidaria para mobilizacao recursos locais culturais naturais or 54000 54000 54000 privada sul pr turvo 853993000144 apoio iniciativa turismo base comunitaria regioao turvo pr ministerio turismo meio departamento qualificacao certificacao producao associada turismo dcpat lancou junho edital chamada publica projetos mtur objetivo selecionar projetos fomento 44155891972 07588185805
292	fomento desenvolvimento turistico local chamada publica programa visa apoiar projetos incubacao

	empreendimentos economicos solidarios cadeia produtiva turismo projetos turismo base principios economia solidaria para mobilizacao recursos locais culturais naturais or 54000 54000 54000 estadual centro oeste mt cuiaba 3507415000900 apoio iniciativa turismo base comunitaria barao melgaco mt ministerio turismo meio departamento qualificacao certificacao producao associada turismo dcpat lancou junho edital chamada publica projetos mtur objetivo selecionar projetos fomento 46881824149 07588185805
298	fomento desenvolvimento turistico local chamada publica programa visa apoiar projetos incubacao empreendimentos economicos solidarios cadeia produtiva turismo projetos turismo base principios economia solidaria para mobilizacao recursos locais culturais naturais or 54000 54000 54000 privada sudeste mg brumadinho 5422243000131 apoio iniciativas turismo base comunitaria brumadinho mg ministerio turismo meio departamento qualificacao certificacao producao associada turismo dcpat lancou junho edital chamada publica projetos mtur objetivo selecionar projetos fomento 74135210678 07588185805

Processamento Sequencial

O algoritmo MaxCluster trabalha na geração de conjuntos frequentes maximais (conjuntos que não são subconjuntos de nenhum outro) e possui os seguintes parâmetros: (i) similaridade mínima aceitável; (ii) tamanho máximo de conjunto. Quanto maior for a similaridade mínima aceitável, mais parecidos devem ser os elementos dos grupos formados. O tamanho máximo de conjunto limita o processamento realizado, podendo da árvore de busca por conjuntos frequentes os grupos que ultrapassam este limiar. Ambos os parâmetros representam um compromisso entre desempenho computacional e qualidade dos grupos gerados. Uma característica do MaxCluster *não desejável* para o problema proposta nesta Fase 1 é a possibilidade de elementos (no caso, convênios) não estarem presentes em nenhum dos grupos formados. Estes comentários servem tanto para a versão sequencial quanto paralela do MaxCluster.

Não se estabeleceu nenhuma restrição em relação à quantidade ideal de grupos gerados na execução. Percebe-se que o tamanho dos grupos não ultrapassa 25 elementos. Uma vez que o algoritmo foi parametrizado com tamanho máximo de grupos acima de 30, este parâmetro não gerou impacto nas execuções. Variou-se então a similaridade mínima para se chegar a resultados que contenham grupos com tamanhos considerados razoáveis em um julgamento inicial, em torno de 13 elementos, maximizando a chance de um convênio qualquer estar presente no resultado. A Figura 9 exibe a variação do número de grupos em função da similaridade mínima aceitável. O resultado selecionado foi o da execução parametrizada com similaridade mínima de 69%.

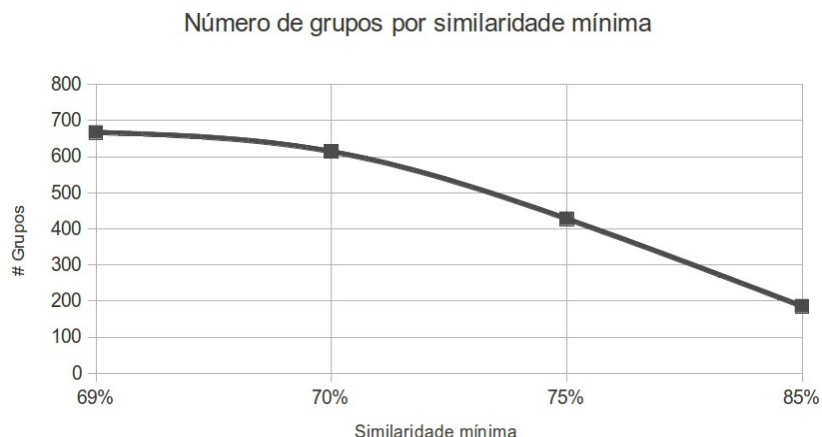


Figura 9. Variação do número de grupos por similaridade mínima no algoritmo MaxCluster.

A Tabela 3 apresenta um trecho do arquivo de saída do algoritmo MaxCluster. Cada linha contém N números, onde os $N-1$ primeiros números identificam os convênios e o N -ésimo número é a similaridade média do grupo.

Tabela 3. Exemplo de agrupamento de convênios

Grupo (id convênio)	Similaridade (%)
{209 292 298 317 394 396 409 413 813 824 880 901 902}	0.754966
{4191 4193 4211 4226 4351 4355 4356 4547 4729 4753 4999 5111 5660 5661 5787 5899 5905 5990 6001 6007 6022 6029 6132 6358}	0.743999
{4226 4351 4547 4732 4751 5061 5097 5102 5103 5111 5789 5897 5921 5938 6132 }	0.771554
{4211 4547 4729 4999 5111 5660 5661 5787 5899 5905 5990 6001 6007 6022 6029 6132 6136}	0.746785
{7767 7775 7780 7821 7872 7874 7953 8053}	0.774280

Agrupamento de Convênios por Similaridade de Empenhos

Esta seção descreve as execuções ligadas ao agrupamento de convênios por similaridade de empenhos, rotulada neste documento de Fase 2.

Preprocessamento

Segue abaixo consulta ao banco de dados realizada para formação da base inicial da Fase 2:

```
select convenios.Nr_Convenio, empenhos.Cd_Natureza_Despesa, empenhos.VI_Nota_Empenho from convenios,empenhos where
convenios.Nr_Convenio=empenhos.Nr_Convenio and convenios.Tx_Situacao like "%Prestação de Contas%" and
convenios.Tx_Modalidade="Convênio" and convenios.Nr_Convenio not in (select distinct(convenios.Nr_Convenio) from convenios,
empenhos where convenios.Nr_Convenio=empenhos.Nr_Convenio and empenhos.Tx_Tipo_Nota_Empenho != "Empenho Original");
```

O resultado da consulta é um conjunto de empenhos de convênios que já foram executados e que não tenham em seu histórico empenhos diferentes dos do tipo “Empenho Original”. Esta consulta representa a primeira fase do fluxo de trabalho definido no PDI, exibido na Figura 10, gerando ao final um arquivo CSV com separador ponto e vírgula, cabeçalho incluso, codificação UTF-8, formato de terminação de linha Unix LF e opção “*Fast data dump (no formatting)*” habilitada.

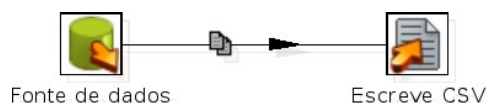


Figura 10. Preprocessamento inicial da Fase 2 realizado no PDI.

O arquivo resultante apresenta o seguinte formato: <id_convênio>, <valor_total_nat_desp_1>, <valor_total_nat_desp_2>, ..., <valor_total_nat_desp_K> (para K diferentes códigos de natureza de despesa). Note-se que cada dimensão do conjunto de dados é um código de natureza de despesa e o valor correspondente ao campo é uma agregação (soma) dos valores de todas as ocorrências de empenhos com aquela natureza de despesa. O conjunto de dados resultantes é uma matriz esparsa (438 atributos), a ser processada pelo algoritmo K-means.

Processamento

A implementação do algoritmo K-means utilizada foi a disponível na ferramenta Weka (<http://www.cs.waikato.ac.nz/ml/weka/>). Foi utilizado no algoritmo um número de conjuntos igual a 667 e no máximo 500 interações. O número de grupos parametrizado foi o mesmo número de grupos encontrados na Fase 1; este casamento é importante para não haver desbalanceamento na Fase 3. A qualidade do agrupamento foi inferida a partir do erro de classificação ao se utilizar o agrupamento em um processo de classificação. O classificador utilizado foi uma árvore de decisão J48 (também uma implementação do Weka), com erro inferido a partir de validação cruzada (*10-fold cross-validation*). O acerto do modelo de classificação foi de 91,91% e pode ser considerado satisfatório, indicando assim, que o agrupamento realizado pelo algoritmo k-means foi aceitável.

Análise de Grupos

A análise de grupos deverá observar os grupos de convênios obtidos na Fase 1, os grupos de convênios obtidos da Fase 2 e produzir os seguintes resultados:

- ✦ Matriz de similaridades de convênios – a similaridade (coeficiente de Jaccard) entre grupos da Fase 1 e grupos da Fase 2, em ordem decrescente;
- ✦ Convênios dissimilares em relação ao empenho – aqueles convênios que estavam no mesmo grupo na Fase 1, mas em grupos diferentes na Fase 2;

- ▲ Convênios dissimilares textualmente – aqueles convênios que estavam no mesmo grupo na Fase 2, mas em grupos diferentes na Fase 1.

A similaridade máxima encontrada na análise foi de 50%. Após algumas execuções, chegou-se a um valor de similaridade mínima de 20% para produzir os resultados (definido *ad hoc*). Este valor resultou em 49 pares de grupos analisados, exibidos na Tabela 4 em ordem decrescente.

Tabela 4. Exemplo de grupos similares.

similaridade[101][368] =	0.5	similaridade[497][594] =	0.2
similaridade[2][576] =	0.333333	similaridade[469][586] =	0.2
similaridade[3][562] =	0.333333	similaridade[53][600] =	0.2
similaridade[242][209] =	0.333333	similaridade[464][544] =	0.2
similaridade[28][269] =	0.333333	similaridade[463][552] =	0.2
similaridade[194][551] =	0.333333	similaridade[64][593] =	0.2
similaridade[391][577] =	0.333333	similaridade[89][320] =	0.2
similaridade[103][518] =	0.333333	similaridade[460][544] =	0.2
similaridade[322][29] =	0.25	similaridade[111][169] =	0.2
similaridade[194][21] =	0.25	similaridade[120][593] =	0.2
similaridade[183][617] =	0.25	similaridade[420][578] =	0.2
similaridade[171][540] =	0.25	similaridade[183][82] =	0.2
similaridade[64][132] =	0.25	similaridade[382][544] =	0.2

similaridade[58][616] = 0.25	similaridade[369][234] = 0.2
similaridade[45][528] = 0.25	similaridade[214][550] = 0.2
similaridade[40][594] = 0.25	similaridade[231][67] = 0.2
similaridade[9][11] = 0.25	similaridade[354][234] = 0.2
similaridade[1][418] = 0.25	similaridade[264][584] = 0.2
similaridade[544][86] = 0.25	similaridade[279][454] = 0.2
similaridade[543][86] = 0.25	similaridade[280][428] = 0.2
similaridade[524][55] = 0.25	similaridade[287][588] = 0.2
similaridade[490][562] = 0.25	similaridade[293][65] = 0.2
similaridade[486][562] = 0.25	similaridade[301][584] = 0.2
similaridade[400][179] = 0.25	similaridade[325][616] = 0.2
similaridade[451][0] = 0.25	

A similaridade mínima é o limiar que determina se pares de grupos analisados devem ser descartados ou não (eles serão descartados se o coeficiente de Jaccard for inferior à similaridade mínima especificada). Com este valor mínimo de similaridade chegou-se em um total de 88 convênios com dissimilaridades em termos de empenhos, e 55 convênios com dissimilaridades textuais. Esses grupos devem ser analisados por especialistas pois representam práticas de convênios. Os mesmos devem apresentar juízo de valor nessas práticas.

Piloto 2 - Geração de Assinaturas de Propostas de Convênios

Agrupamento de Convênios por Campos Textuais

A solução inicial para geração de palavras-chave inicia-se pelo agrupamento dos convênios considerados similares. Optou-se pela utilização de agrupamentos baseado em texto utilizando o campo justificativa do convênio. Almeja-se que o agrupamento textual seja capaz de utilizar a riqueza semântica existente nas descrições textuais para formar grupos altamente similares. A Tabela 1 apresentou um exemplo do texto utilizado.

Aplicando uma similaridade mínima de 0.69 e determinando que cada conjunto tenha no máximo 50 elementos um total de 667 grupos dos quais 169 grupos de convenio foram gerados. A Tabela 2 apresenta um exemplo dos agrupamentos realizados pelo algoritmo *MaxCluster*. A justificativa dos três primeiros convênios (209, 292 e 298) do primeiro grupo exibidos na Tabela 2 foram apresentados na Tabela 1. Nota-se a similaridade entre as justificativas dos convênios. Observe ainda, que o referido grupo apresenta uma similaridade de aproximadamente 0.75 (75%).

Após o processo de agrupamento realizou-se a rotulação de cada grupo. Esse procedimento é aplicado para permitir a posterior classificação de uma nova proposta de convênio. A Figura 11 apresenta um exemplo ilustrativo de todo processo.

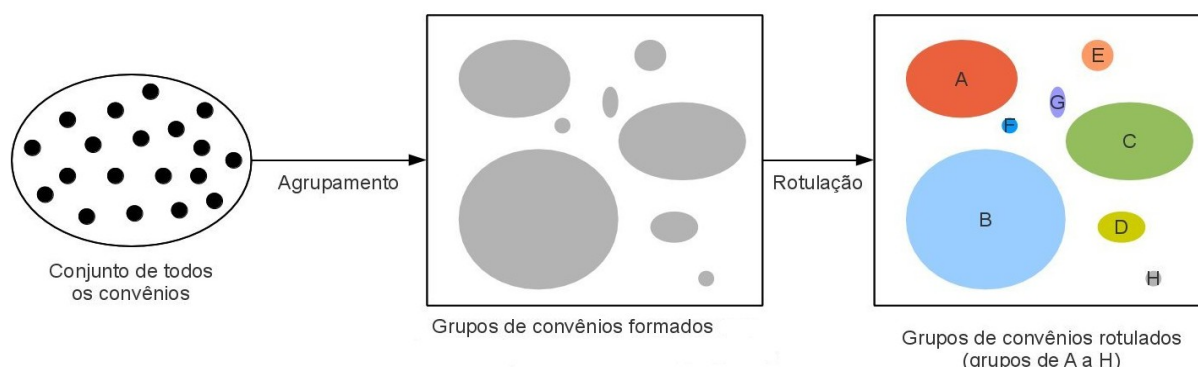


Figura 11. Exemplo ilustrativo do processo de agrupamento e rotulação.

A partir do conjunto de convênios agrupados e rotulados, pode-se extrair de cada grupo um conjunto de palavras chave. Ao cadastrar um novo convênio, por exemplo, com características similares ao grupo A, pode-se então sugerir as palavras chave que representam esse grupo. Para selecionar o conjunto de palavras chave de cada grupo aplicou-se uma análise local de cada grupo. Selecionou-se as K palavras mais frequentes de cada grupo; inicialmente, aplicou-se $k=10$. Esse parâmetro deve ser ajustado com o intuito de fornecer um número apropriado de palavras chave para cada grupo.

Observe na Tabela 5 que algumas palavras chave obtidas foram valores numéricos e palavras pobres semanticamente, por exemplo, “para”. Essas características apontam a necessidade de remover essas palavras antes do processo de agrupamento e a necessidade de uma estratégia mais elaborada para a seleção das palavras chave de cada grupo formado. Uma solução interessante é integrar as informações obtidas de cada agrupamento com um vocabulário controlado de governo eletrônico.

Tabela 5. Conjunto de palavras-chave geradas para os agrupamentos com rótulos de 1 a 6.

Rotulo	Conjunto de palavras-chave
1	{para 24000 centro computadores municipio populacao desenvolvimento programa tecnico}
2	{saude 36000 atencao para medicamentos aquisicao basica farmaceutica assistencia insumos}
3	{produtores social para sub adequacao 53000 22000 54000 garantia programa}
4	{para saude 36000 atencao medicamentos municipal basica municipio aquisicao estado}
5	{para saude 36000 municipio medicamentos atencao apoio basica municipal farmaceutica}
6	{saude 36000 para atencao medicamentos basica assistencia municipal aquisicao farmaceutica}

Classificando uma Nova Proposta

A sugestão de palavras chave para uma nova proposta de convênio demanda um processo de classificação. Em outras palavras, para um nova proposta é necessário determinar de qual grupo ela mais se aproxima. De posse dessa informação, para uniformizar o cadastro de propostas de convênios, pode-se

sugerir o conjunto de palavras chave do agrupamento similar. A Figura 12 apresenta uma ilustração do processo de classificação empregado. Observe que o classificador utiliza as informações provenientes do processo de agrupamento para gerar um modelo. Quando uma nova proposta é apresentada a probabilidade dessa proposta pertencer a cada agrupamento/classe é exibida. A maior probabilidade ou as maiores representam a(s) classe(s) mais provável(eis) do novo convênio.

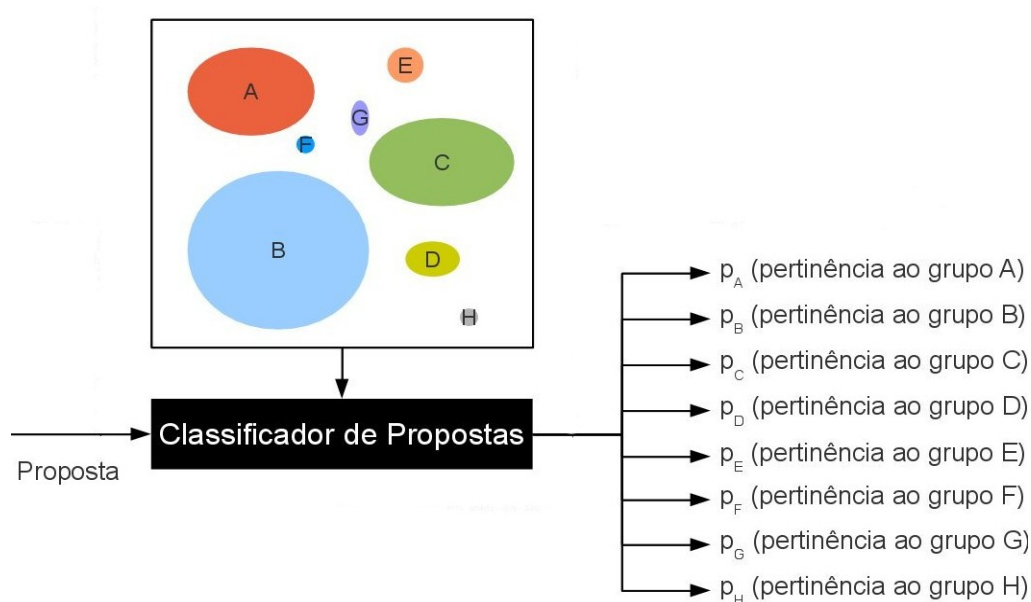


Figura 12. Exemplo ilustrativo do processo de classificação.

Dentre os diversos paradigmas de classificação existentes na literatura aplicou-se um classificador *lazy*. Esse tipo de classificador foi escolhido por gerar um modelo de forma mais rápida e sob demanda; evitando assim, um processo longo de treinamento. Foi aplicado o algoritmo LAC (*lazy associative classification*), proveniente do convênio SERPRO-UFMG.

O texto da justificativa de cada convênio foi transformado em um exemplo para o classificador. A Tabela 6 apresenta exemplos de instâncias utilizadas no processo de treinamento e classificação. Note que o rótulo de cada grupo origina uma classe; assim, obtêm-se um total de 169 classes inicialmente. Já o texto utilizado na justificativa do convênio representam os atributos.

Tabela 6. Exemplo de instâncias utilizadas no processo de treinamento e classificação.

Id	Instância e classe
155	CLASS=1 w=ae w=promocao w=assistencia w=farmaceutica w=insumos w=estrategicos w=atencao w=basica w=saude w=estado w=rio w=grande w=sul w=emenda w=parlamentar w=apoio w=assistencia w=farmaceutica w=ambito w=atencao w=basica w=saude w=bem w=medicamentos w=para w=programas w=saude w=especificos w=cuidados w=nivel w=atencao w=36000 w=36000 w=36000 w=municipal w=sul w=rs w=imigrante w=92454776000108 w=aquisicao w=medicamentos w=municipio w=imigrante w=populacao w=habitantes w=conforme w=censo w=ibge w=necessita w=contar w=apoio w=governo w=federal w=para w=atender w=demandas w=populacao w=projeto w=destina w=aquisicao w=medicamentos w=uso w=hospitalar w=ambulatoria w=28900456091 w=05985781100
156	CLASS=1 w=ae w=promocao w=assistencia w=farmaceutica w=insumos w=estrategicos w=atencao w=basica w=saude w=estado w=rio w=grande w=sul w=emenda w=parlamentar w=apoio w=assistencia w=farmaceutica w=ambito w=atencao w=basica w=saude w=bem w=medicamentos w=para w=programas w=saude w=especificos w=cuidados w=nivel w=atencao w=36000 w=36000 w=36000 w=municipal w=sul w=rs w=chapada w=87613220000179 w=aquisicao w=medicamentos w=municipio w=chapada w=populacao w=habitantes w=conforme w=censo w=ibge w=necessita w=contar w=apoio w=governo w=federal w=para w=atender w=demandas w=populacao w=projeto w=destina w=aquisicao w=medicamentos w=uso w=hospitalar w=ambulat w=37319353091 w=05985781100

Piloto 3 - Análise de Padrões em Grafos de Convênios

Este piloto visa agregar valor às análises de padrões formados por grafos baseados em dados de convênios. A seguinte estrutura de grafo foi projetada: grafos em que vértices são convênios (rotulados pelos seus atributos) e arestas são relacionamentos estabelecidos por agrupamento prévio de convênios (arestas não direcionadas e não rotuladas). Os padrões esperados para a estrutura de grafo são subgrafos densos considerando apenas atributos de convênios, utilizando o conhecimento *a priori* de agrupamento entre convênios para definir conectividade. Para o processo de mineração foi empregado o algoritmo SCPM (*Structural Correlation Pattern Mining*) proveniente do convênio Serpro/UFMG.

Agrupamento de Convênios por Similaridade Textual (Fase 1)

A primeira fase deste piloto empregou o mesmo procedimento adotado na fase 1 do piloto 1.

Descoberta de Subgrafos Densos (Fase 2)

Nesta segunda fase, os grupos encontrados na Fase 1 são usados para criação de um grafo de convênios, onde convênios são vértices e para todo par de convênios (i, j) existe uma aresta entre eles não

direcionada e não rotulada se, e somente se, i e j pertencem ao mesmo grupo. Ainda nesta fase, o grafo formado é minerado para descoberta de subgrafos densos, onde a densidade é definida a partir das informações de conectividade entre vértices e frequência de coocorrência de valores de seus atributos. O processo completo é ilustrado na Figura 13.

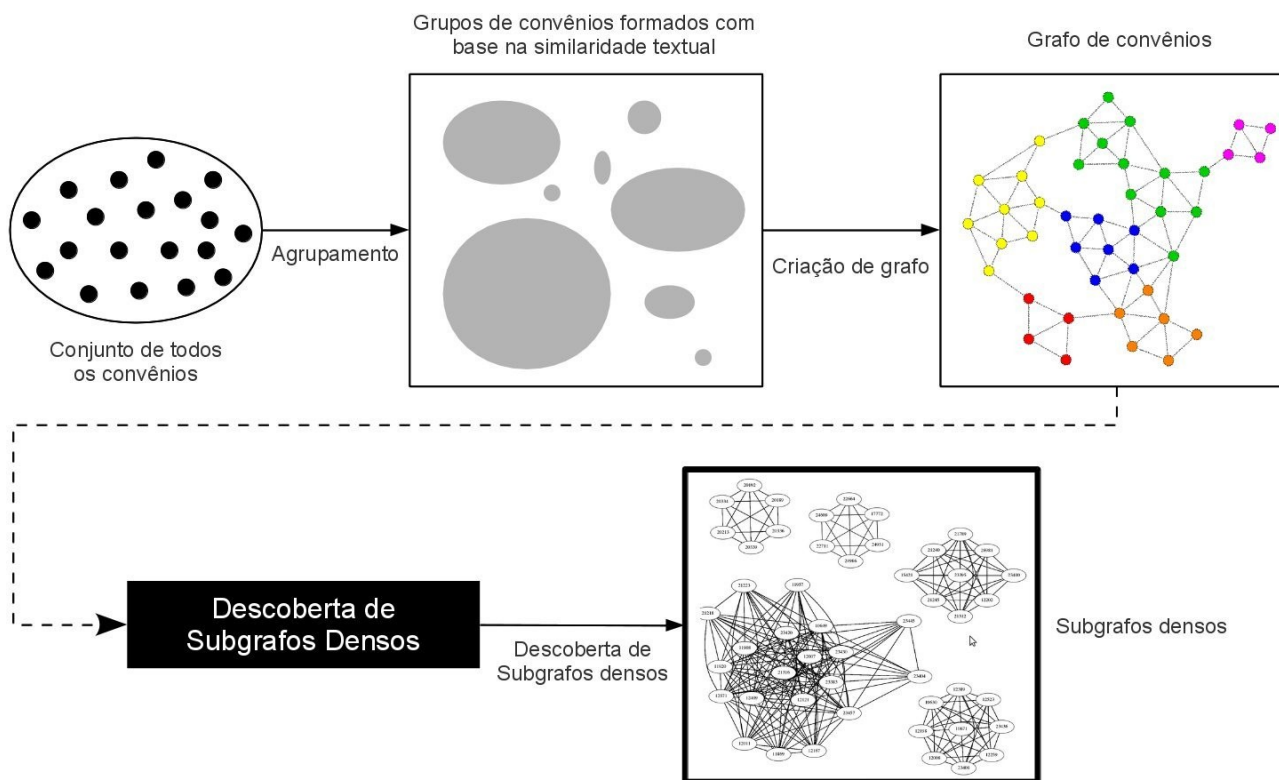


Figura 13. Descoberta de subgrafos densos

Preprocessamento

A primeira rotina de preprocessamento executada foi a geração do arquivo de vértices que descreve o grafo de convênios. O arquivo de vértices é formado por uma lista com o identificador de cada um dos vértices do grafo seguido pelo identificador de todos os seus vizinhos. A Tabela 7 apresenta parte do arquivo de vértices utilizado.

Tabela 7. Exemplo de arquivo de vértices.

```
04203,26857
04221,03197,03208,03284
04248,04251,04252,04254,04257,04259,04263,04642,04755,04802,04883
04251,04248,04252,04254,04257,04259,04263,04281,04282,04594,04642,04755,04882,04883
04252,04248,04251,04254,04257,04259,04263,04281,04282,04594,04642,04755,04882,04883
04254,04248,04251,04252,04257,04259,04263,04281,04282,04594,04642,04755,04882,04883
04257,04248,04251,04252,04254,04259,04263,04281,04282,04594,04642,04755,04882,04883
04259,04248,04251,04252,04254,04257,04263,04281,04282,04594,04642,04755,04882,04883
```

```
04263,04248,04251,04252,04254,04257,04259,04281,04282,04594,04642,04755,04882,04883
04276,03721,04408,04543,04672,15268
04281,04251,04252,04254,04257,04259,04263,04282,04594,04642,04755,04882,04883
04282,04251,04252,04254,04257,04259,04263,04281,04594,04642,04755,04882,04883
```

O segundo passo de pré-processamento foi gerar o arquivo CSV de atributos de convênios. Segue abaixo consulta ao banco de dados realizada para formação da base de atributos de vértices do grafo (convênios). O resultado da consulta são valores de atributos de convênios que já foram executados e que não tenham em seu histórico empenhos diferentes dos do tipo “Empenho Original”.

```
select convenios.Ano_Proposta, convenios.Cd_Programa, convenios.VI_Repasse, convenios.VI_Contrapartida_Total,
convenios.Qt_Empenhos, convenios.Qt_Aditivos, convenios.Qt_Prorroga, convenios.Tx_Qualific_Proponente,
convenios.Id_Prop_Programa from convenios where convenios.Tx_Situacao like "%Prestação de Contas%" and
convenios.Tx_Modalidade="Convênio" and convenios.Nr_Convenio not in (select distinct(convenios.Nr_Convenio) from convenios,
empenhos where convenios.Nr_Convenio=empenhos.Nr_Convenio and empenhos.Tx_Tipo_Nota_Empenho != "Empenho Original");
```

Os arquivos vertices.csv e atributos.csv gerados no pré-processamento serão a entrada no algoritmo SCPM.

Processamento

O SCPM foi executado com os seguintes argumentos: `./scpm -s 100 -g 0.5 -k -1 -q 5 -o output.csv -a atributos.csv -n vertices.csv`. Os arquivos atributos.csv e vertices.csv são os arquivos de atributos e vértices, respectivamente, enquanto o arquivo output.csv é a saída do algoritmo. A opção “-s” permite especificar o suporte mínimo, “-g” a densidade mínima dos *quasi*-cliques, “-k” é o número máximo de subgrafos encontrados e “-q” é o tamanho mínimo dos *quasi*-cliques.

A tabela 8 apresenta um trecho do arquivo output.csv. A primeira linha exibe as características do atributo escolhido, na segunda linha todos os vértices que entraram na análise e nas linhas subsequentes os subgrafos densos encontrados. Para visualização dos grafos aplicou-se a biblioteca Graphviz (<http://www.graphviz.org/>).

Tabela 8. Exemplo de saída do algoritmo SCPM.

```
attribute_set=Convênio,size_attribute_set=1,support=27145,size_coverage=16,epsilon=0.000589427,delta=0.655411,expected_epsilon=0.000899325
coverage=11920,15421,21218,21223,21240,21245,21312,21789,23383,23395,23404,23410,23420,23430,23445
attribute_set=Convênio,size=5,gamma=0.5,vertices=11920-21218-21223-23383-23430
attribute_set=Convênio,size=5,gamma=0.5,vertices=23404-23420-23430-23445-23457
```

A Figura 14 apresenta exemplos de subgrafos densos extraídos. Novamente, cabe ao especialista avaliar a qualidade dos grafos gerados.

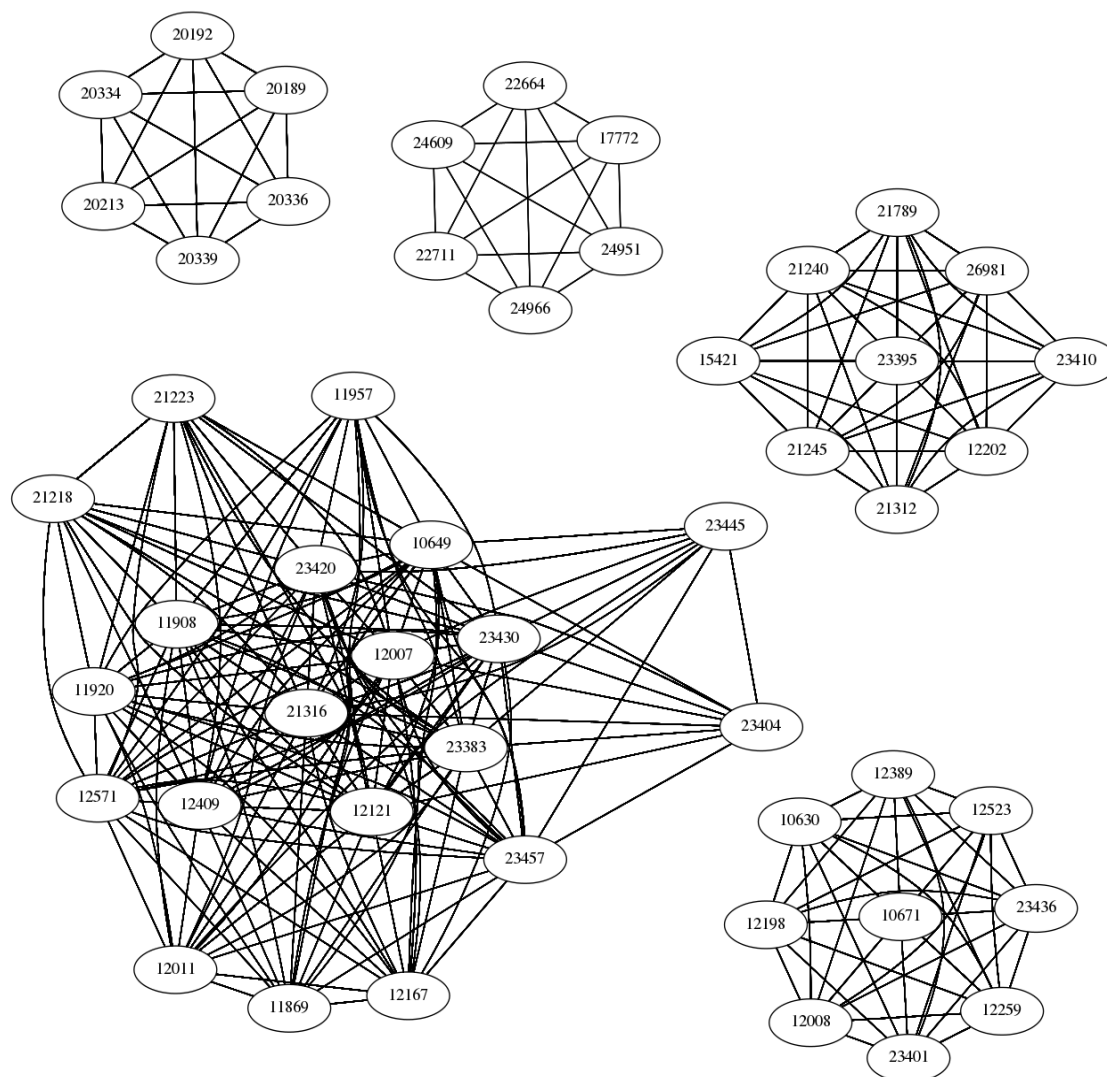


Figura 14. Subgrafos densos encontrados.



UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG
COORDENADORIA DE TRANSFERÊNCIA E INOVAÇÃO TECNOLÓGICA – CTIT
PRÓ - REITORIA DE PESQUISA

SERVIÇO FEDERAL DE PROCESSAMENTO DE DADOS – SERPRO

EDITAL Nº 01/2013

ESTE EDITAL SE REGERÁ PELA LEI Nº 10.973/2004, DECRETO Nº 5.563/2005 E LEI 8.666/93.

01 – PREÂMBULO

A UNIVERSIDADE FEDERAL DE MINAS GERAIS – UFMG, por meio de sua **PRÓ- REITORIA DE PESQUISA/ COORDENADORIA DE TRANSFERÊNCIA E INOVAÇÃO TECNOLÓGICA – CTIT** e o **SERVIÇO FEDERAL DE PROCESSAMENTO DE DADOS – SERPRO** tornam público e comunicam aos interessados que, até o dia 10 de janeiro de 2014, às 17:00 horas, na Avenida Antônio Carlos, nº 6627, – Unidade Administrativa II - 2º andar, - sala 2008/2017, Bairro Pampulha, em Belo Horizonte, Minas Gerais, Brasil, CEP 31.270.901, receberá os envelopes dos interessados contendo a documentação prevista no item 06 do presente Edital.

02 - DO OBJETO

2.1- O presente Edital tem por objetivo a seleção de proposta mais vantajosa para a contratação de empresa, ou consórcio de empresas, para a transferência, a título oneroso, sem exclusividade, no Brasil e no exterior, dos direitos para **uso, produção, comercialização, prestação de serviços**, relacionados ao *Know-How* intitulado “**Práticas de Mineração de Dados Escalável para Sistemas de Governo**”, de titularidade da UFMG e SERPRO, descrito no anexo I.



03 – DA DISPENSA DE LICITAÇÃO

3.1- É dispensável, nos termos do art. 24, inciso XXV, da Lei n.º 8.666, de 21 de junho de 1993, a realização de licitação em contratação realizada por Instituição Científica e Tecnológica - ICT para a transferência de tecnologia e para o licenciamento de direito de uso ou de exploração de criação protegida.

04- DA DESCRIÇÃO SUCINTA DO *KNOW-HOW*

O *Know-how* trata-se de conjunto de algoritmos paralelos para mineração de grandes volumes de dados, a plataforma para paralelização de algoritmos Anthill e uma proposta de plataforma integrada para serviços em mineração de dados. Ele engloba todo conhecimento adquirido na aplicação de técnicas de mineração de dados escalável em dados de sistemas governamentais. Esse *know-how* reúne o conhecimento acadêmico em mineração de dados escalável e o conhecimento em governo eletrônico com o intuito de prover soluções para problemas relacionados ao governo. A junção desses conhecimentos resulta em especialistas capazes de definir aplicações, modelos, algoritmos e sistemas.

Especificamente, este *Know-how* engloba os seguintes conhecimentos: experimentação de técnicas de mineração de dados escalável em bases de dados de governo, novo algoritmo para agrupamento de texto (MaxCluster) e a aplicação das técnicas em três problemas reais: agrupamento de texto, geração de assinaturas textuais e mineração de grafos.

05 - DA ENTREGA DOS ENVELOPES

5.1- Até a data, horário e local indicados no preâmbulo deste Edital, os interessados deverão protocolizar o envelope devidamente lacrado e indevassado, com a documentação prevista no item 06 deste Edital, contendo em sua parte externa e fronteira, além da razão social da proponente, os seguintes dizeres:

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG

EDITAL Nº 01/2013 - PARA A TRANSFERÊNCIA, A TÍTULO ONEROSO, SEM EXCLUSIVIDADE, NO BRASIL E NO EXTERIOR DOS DIREITOS, PARA O USO, PRODUÇÃO, COMERCIALIZAÇÃO, PRESTAÇÃO DE SERVIÇOS, RELACIONADOS AO *KNOW-HOW* INTITULADO “PRÁTICAS DE MINERAÇÃO DE DADOS ESCALÁVEL PARA SISTEMAS DE GOVERNO”.

Avenida Antônio Carlos, 6627- Unidade Administrativa II - 2º andar

Sala 2008/2017, Bairro Pampulha



Belo Horizonte, Minas Gerais - CEP 31.270.901

5.2- Será admitida documentação enviada pelo Correio, desde que entregue à Universidade até a data, horário e local indicados no preâmbulo deste Edital, sendo de inteira responsabilidade do proponente os riscos pelo envio da documentação desta forma.

5.3 - Não serão admitidas quaisquer retificações na documentação apresentada.

5.4 - Os envelopes entregues em local diferente ou dias e horários posteriores aos especificados não serão objeto de análise pelo Comitê Técnico.

5.5- As propostas, que serão formadas pelas declarações e relatórios exigidos no item 7 deste Edital, deverão ser impressas com tinta indelével e assinadas pelo representante legal da empresa proponente, autorizado a contrair obrigações em seu nome, devidamente identificado. Todas as páginas da proposta deverão ser seqüencialmente numeradas e rubricadas pelo signatário da proposta.

5.6 - As propostas não poderão conter rasuras, emendas ou entrelinhas que obscureçam seu perfeito entendimento e não serão aceitas propostas enviadas por telex, fax, telegrama ou via Internet.

06- DA DOCUMENTAÇÃO

6.1 - Os envelopes deverão conter:

6.1.1 - Comprovação da Regularidade Jurídica, Fiscal e Trabalhista

- Ato Constitutivo, Estatuto ou Contrato Social, em vigor, devidamente registrado, em se tratando de sociedades empresárias/simples, e, no caso de sociedades empresárias por ações, acompanhado de documentos de eleição de seus administradores.

- Certidão de Regularidade do Fundo de Garantia por Tempo de Serviço (FGTS) - CRF, administrado pela Caixa Econômica Federal, com prazo de validade em vigor.

- Certidão Negativa de Débito com o Instituto Nacional do Seguro Social – INSS, com prazo de validade em vigor.

**Avenida Antônio Carlos 6627 - Unidade Administrativa II, 2º andar, sala 2017, CEP 30.161-970 – Belo Horizonte – MG, Fone: (31) 3409-4033, 3409-6465, Telefax: 3409-6430-E-mail: info@ctit.ufmg.br
<http://www.ufmg.br/ctit>**



UFMG

- Certidão Conjunta Negativa de Débitos Relativos a Tributos Federais e à Dívida Ativa da União, com prazo de validade em vigor.
- Certidão Negativa de Débitos Trabalhistas.
- Prova de regularidade para com as Fazendas Estadual e Municipal do domicílio ou sede da empresa participante, com prazo de validade em vigor.
- Certidão negativa de falência e/ou concordata/ recuperação judicial emitida até 90 (noventa) dias anteriores à data indicada no preâmbulo deste Edital, se a referida certidão não contiver data de validade.

6.1.2 – Comprovação da Qualificação Técnica e Econômico - Financeira para exploração das tecnologias

- Atividade econômica e histórico da empresa condizente com o objeto a ser licenciado, por meio de declaração do representante legal da proponente;
- Balanço patrimonial e demonstrações contábeis do último exercício social, já exigíveis e apresentados na forma da lei, que comprovem a boa situação financeira da empresa, vedada a sua substituição por balancetes ou balanços provisórios, podendo ser atualizados por índices oficiais quando encerrados há mais de 3 (três) meses da data de apresentação da proposta.

α) As empresas que não tiverem concluído um exercício social poderão apresentar o balanço de abertura;

β) Entende-se por "apresentados na forma da lei" balanço patrimonial e demonstrações contábeis que deverão ser apresentados em cópias autenticadas das folhas do Livro Diário do Proponente (onde esses dados estão registrados), devidamente registrado no Órgão competente, acompanhados de cópias autenticadas dos termos de abertura e encerramento do respectivo Livro, estes últimos assinadas pelo proprietário ou por quem de direito e por profissional de contabilidade legalmente habilitado,



UFMG

χ) As sociedades anônimas que sejam obrigadas por lei a publicar seus resultados contábeis poderão, em substituição às cópias do Livro diário, apresentar cópia (autenticada por cartório ou servidor público) do balanço patrimonial publicado em “diário oficial”.

δ) A comprovação da boa situação financeira da empresa será baseada na obtenção de índices maiores que 1 (um) relativamente a Liquidez Geral (LG), Solvência Geral (SG) e Liquidez Corrente (LC), resultantes da aplicação das fórmulas:

$$\begin{aligned} & \text{Ativo Circulante} + \text{Realizável a Longo Prazo} \\ \text{LG} = & \frac{\text{Passivo Circulante} + \text{Exigível a Longo Prazo}}{\text{Ativo Total}} \\ \text{SG} = & \frac{\text{Passivo Circulante} + \text{Exigível a Longo Prazo}}{\text{Ativo Circulante}} \\ \text{LC} = & \frac{\text{Passivo Circulante}}{\text{Passivo Circulante}} \end{aligned}$$

6.1.3- Em se tratando de Consórcio de empresas, cada consorciado deverá atender a todas as condições relativas à documentação jurídica, fiscal, trabalhista, técnica e econômico-financeira, devendo apresentar ainda:

1) Comprovação do compromisso público ou particular de constituição do consórcio subscrito pelos consorciados, contendo:

- a designação da empresa que representará o consórcio;
- as empresas integrantes do consórcio;
- o objetivo da consorciação;
- os compromissos e obrigações dos consorciados.

2) Declaração expressa de responsabilidade solidária de todos os consorciados pelos atos praticados sob o consórcio, em relação à eventual contratação



UFMG

3) Termo de Compromisso firmado pelos consorciados de que o consórcio não terá sua composição ou constituição alteradas, ou sob qualquer forma modificadas, sem prévia e expressa anuência, escrita da UFMG e do SERPRO, durante a vigência da contratação decorrente da presente oferta pública.

6.1.3.1- Quanto às exigências relativas à qualificação econômico-financeira, admitir-se-á o somatório dos valores de cada consorciado, na proporção de sua respectiva participação.

6.1.3.2- Na hipótese de o detentor da melhor proposta ser consórcio, deverá apresentar, antes da contratação, a documentação comprobatória de sua constituição e registro, nos termos do compromisso assumido para participar da oferta pública objeto do presente Edital.

6.1.3.3- A empresa representante do consórcio deverá atender as condições de liderança, cabendo obrigatoriamente, à empresa brasileira, em se tratando de consórcios de empresas brasileiras e estrangeiras.

6.1.4- - As empresas estrangeiras que não funcionem no país, tanto quanto possível, atenderão a todas as exigências constantes deste item 6, mediante documentos equivalentes, autenticados pelos respectivos consulados e traduzidos por tradutor juramentado, devendo ter representação legal no Brasil com poderes expressos para receber citação e responder administrativa ou judicialmente.

6.1.5- Proposta Comercial, conforme subitens 5.5 e 5.6 deste edital.

6.1.6- Fica o proponente obrigado a declarar, sob as penalidades cabíveis, relativamente aos documentos exigidos neste Edital, se existir após a data de sua emissão, algum fato superveniente que impeça a contratação.

07- CRITÉRIOS TÉCNICOS OBJETIVOS PARA SELEÇÃO DA CONTRATAÇÃO MAIS VANTAJOSA

7.1 - Critério Comprovação Pontos.

n.º	CRITÉRIO	COMPROVAÇÃO	PONTOS (1-10)
1	Possuir setor/departamento de P&D em atividade. No caso de consórcio, será considerado aquele da empresa onde será desenvolvido o <i>Know-How</i> .	Declaração da empresa.	10 pontos



UFMG

2	Possuir equipe técnica para desenvolvimento e aplicação do <i>Know-How</i> . No caso de consórcio, será considerada a equipe da empresa onde será desenvolvido o produto.	Comprovação do vínculo profissional dos membros da equipe com a empresa envolvida no desenvolvimento do produto e cópia dos certificados de titulação.	Doutor /3 pontos por membro Mestre /2 pontos por membro Graduado /Especialista /1 pontos por membro
3	Possuir infra-estrutura técnico-operacional para a exploração do <i>Know-How</i> , a exemplo de equipamentos e plataformas próprios ou terceirizados. No caso de terceirizados deverá garantir a infra-estrutura com apresentação do contrato. No caso de consórcio, será considerado a infra-estrutura da empresa onde será desenvolvido o produto.	Declaração da empresa.	10 pontos
4	Ter experiência de pesquisa e desenvolvimento em parceria com instituição de pesquisa no Brasil. No caso de consórcio, serão consideradas as parcerias da empresa onde o <i>Know-How</i> será explorado.	Descrição sucinta dos projetos comprovada por meio de declaração da empresa.	5 pontos para cada projeto comprovado
5	Ter lançado novo produto no último ano. No caso de consórcio, serão considerados os produtos da empresa onde será desenvolvido o produto.	Relatório contendo a data de lançamento e as informações do produto.	5 pontos para cada produto lançado
6	Demonstrar valor estimado de faturamento líquido anual em razão da exploração do <i>Know-How</i> , durante o prazo de vigência do contrato.	Relatório técnico.	Classificação de 1 a 10 pontos, sendo que a pontuação máxima será atribuída para a proposta que apresentar o maior valor de faturamento líquido e a pontuação mínima para a proposta que apresentar menor valor de faturamento líquido.

7	Indicar prazo viável para o início da exploração comercial do <i>Know-How</i> .	Declaração da empresa.	Classificação de 1 a 5 pontos para cada etapa, sendo que a pontuação máxima será atribuída para a proposta que apresentar o menor prazo viável para cada etapa e pontuação mínima para a proposta que apresentar o maior prazo viável para cada etapa.
8	Proposta de remuneração inicial para acesso ao <i>Know-How</i> a ser paga conjuntamente à UFMG e ao SERPRO de no mínimo R\$ 5.000 (cinco mil reais). O valor a ser pago a título de remuneração inicial será dividido pela UFMG e pelo SERPRO na proporção de 50%(cinquenta por cento) para cada.	Declaração da empresa.	Classificação de 1 a 10 pontos, sendo que a pontuação máxima será atribuída para a proposta que apresentar maior remuneração inicial e a pontuação mínima para a proposta que apresentar a menor remuneração inicial.
9	Proposta de percentual de <i>royalties</i> que incidirá sobre o faturamento líquido semestral (= valor bruto auferido na comercialização, deduzidos os tributos incidentes) resultantes da exploração comercial do <i>Know-How</i>, durante o prazo de vigência do contrato. Parâmetros: Percentual de no mínimo: 8% (oito por cento)	Declaração da empresa	Classificação de 1 a 10 pontos, sendo que a pontuação máxima será atribuída para a proposta mais vantajosa e a pontuação mínima para a proposta menos vantajosa.
10	Preferência na realização de parceria com a UFMG e SERPRO para o aprimoramento do <i>Know-How</i> , para a avaliação da versão β (beta) da plataforma em softwares livres disponibilizados, com base no <i>Know-How</i> .	Declaração da empresa	5 pontos



7.2 - Caso mais de uma empresa do consórcio venha a participar do desenvolvimento do *Know-How*, será exigida, na comprovação dos critérios elencados no presente item, a indicação da proporção da sua participação, observado o disposto no item 6.1.3.

7.2.1 – Para efeito de qualificação técnica e econômica – financeira esta participação será considerada pelo Comitê Técnico, desde que seja relevante para aferição do critério, observado o disposto no item 6.1.3.

08 – DA ABERTURA DOS ENVELOPES E SELEÇÃO

8.1 - A seleção das propostas será feita por um Comitê Técnico designado pela Pró-Reitoria de Pesquisa/Coordenadoria de Transferência e Inovação Tecnológica – CTIT e SERPRO.

8.2- Após a abertura dos envelopes contendo a documentação, toda ela será devidamente rubricada pelos membros do Comitê Técnico.

8.3- Será vencedora a proposta cujo somatório da pontuação de todos os critérios perfaça a maior pontuação total.

8.4- O cálculo da pontuação referente aos critérios 6); 7), 8) e 9) do item 7 deste Edital, será feito atribuindo a pontuação máxima à proposta mais vantajosa para a **UFMG** e o **SERPRO** a pontuação mínima para a proposta menos vantajosa. A pontuação das propostas intermediárias será calculada por meio de interpolação linear.

8.5- Na seleção das propostas, o Comitê Técnico poderá, a seu critério, solicitar assessoramento técnico a órgãos ou a profissionais especializados.

8.6- A **UFMG** e o **SERPRO** darão preferência à contratação de empresa de pequeno porte e microempresas, em caso de igualdade de condições, de forma a atender ao disposto no art. 7º, inciso V, § 3º do Decreto 5.563/05.



8.7- Em caso de empate entre as propostas, o Comitê Técnico considerará isoladamente cada item, respeitada a ordem abaixo estabelecida para o desempate e classificação.

8.7.1 – Empresa de Pequeno Porte e microempresas;

8.7.2 – Menor prazo apresentado para comercialização da tecnologia;

8.7.3 - Maior valor apresentando para remuneração inicial;

8.7.4 - Maior percentual apresentado de royalties escalonado.

9- APRESENTAÇÃO DO RESULTADO:

9.1 – O resultado final da seleção da proposta vencedora será publicado no Diário Oficial da União - DOU e disponibilizado na página eletrônica da CTIT/UFMG <http://www.ufmg.br/ctit> até o dia 30 de janeiro de 2014.

10-ANULAÇÃO OU REVOGAÇÃO DO EDITAL

10.1 - A Administração poderá revogar o presente Edital por razões de interesse público, decorrente de fato superveniente devidamente comprovado, devendo anulá-lo, por ilegalidade, de ofício ou por provocação de terceiros, não cabendo às empresas participantes, qualquer indenização por parte da UFMG e do SERPRO.

11-DOS RECURSOS ADMINISTRATIVOS

11.1- Das decisões do Comitê Técnico caberá recurso no prazo de cinco dias úteis, a contar da data da publicação do resultado da seleção, dirigido ao Pró-Reitor de Pesquisa da UFMG, por intermédio do - Comitê Técnico e protocolizado na CTIT até às 17:00h do quinto dia útil.

12 -LOCAL E HORÁRIO PARA INFORMAÇÕES

12.1- As empresas proponentes poderão solicitar esclarecimentos complementares sobre este edital à CTIT, por escrito, via e-mail (info@ctit.ufmg.br) e por fax, através do nº 55-31-3409-6430, de segunda a sexta-feira, no horário das 08:00 às 17:00 horas, até no prazo máximo de 03 (três) dias úteis que antecederem a data marcada para a entrega dos Envelopes. Todos os esclarecimentos complementares deverão conter o número deste edital, seguido pelo nome da empresa.



UFMG

13 - DOCUMENTOS PARA USO E APLICAÇÃO DO *KNOW-HOW*:

13.1 – Os seguintes documentos serão entregues ao detentor da melhor proposta:

1. “Documento da Arquitetura de Referência”.
2. “Documento de Desenvolvimento”.
3. “Documento de Teste da Plataforma Integrada de Serviços Escalável em Mineração de Dados - PISeMD” (versão alfa).

14- DAS DISPOSIÇÕES GERAIS

14.1- Os participantes arcarão com todos os custos decorrentes da elaboração de suas propostas, não cabendo-lhes qualquer ressarcimento ou indenização.

14.2 - Da abertura dos envelopes e da seleção das propostas serão lavradas atas próprias, assinadas pelos membros do Comitê Técnico.

14.3- Os casos omissos serão resolvidos pelo Comitê Técnico nos termos da legislação vigente.

14.4 – A minuta contratual é parte integrante do presente Edital (Anexo I).

14.5 – No ato de assinatura do contrato, a empresa deverá estar cadastrada no Sistema de Cadastramento Unificado de Fornecedores – SICAF.

14.6 – As empresas interessadas em conhecer o texto integral da descrição do Know-How deverão assinar o termo de confidencialidade, cujo modelo estará disponível na página eletrônica <http://www.ufmg.br/ctit> e encaminhá-lo a CTIT para obtenção da cópia.

15. DA NÃO EXCLUSIVIDADE DA TRANSFERÊNCIA

Considerando que a transferência do Know-how será realizada sem exclusividade para a empresa ou o consórcio de empresas detentor da melhor oferta, a UFMG e o SERPRO poderão, a seu exclusivo critério,



UFMG

selecionar mais de uma proposta apresentada ao presente Edital e não estarão impedidos de realizar futuras transferências sem exclusividade.

A participação dos interessados implica na aceitação integral e irrevogável dos termos deste EDITAL, seu(s) anexo(s) e instrumentos, bem como a observância dos regulamentos, normas e das disposições legais pertinentes.

Belo Horizonte, 28 de novembro de 2013.

Universidade Federal de Minas Gerais – UFMG
Pró-Reitoria de Pesquisa
Prof. Renato de Lima Santos

Serviço Federal de Processamento de Dados - SERPRO
Diretor Presidente
Marcos Vinícius Ferreira Mazoni



UFMG

**CONTRATO DE TRANSFERENCIA DE KNOW
HOW N.º QUE ENTRE SI CELEBRAM
A UNIVERSIDADE FEDERAL DE MINAS
GERAIS - UFMG, O SERVIÇO FEDERAL DE
PROCESSAMENTO DE DADOS – SERPRO,
E , DECORRENTE DO EDITAL DE
OFERTA PUBLICA Nº 01/2013.**

A **UNIVERSIDADE FEDERAL DE MINAS GERAIS**, autarquia federal de regime especial, sediada na Avenida Antônio Carlos, nº 6.627, Belo Horizonte, Minas Gerais, inscrita no CNPJ sob o nº 17.217.985/0001-04, doravante denominada **UFMG**, por meio de sua Pró-Reitoria de Pesquisa, representada por seu Pró-Reitor, **Professor Renato de Lima Santos**, portador da CI 17.883.708-8 expedida por SSP/SP, CPF nº 071.522.838-26, residente e domiciliado em Belo Horizonte, Minas Gerais, o **SERVIÇO FEDERAL DE PROCESSAMENTO DE DADOS**, sediado no SGAN-Quadra 601, Módulo “V”, CEP: 70836-900, Asa Norte, Brasília, Distrito Federal, denominado **SERPRO**, neste ato representado por seu Diretor Presidente, Sr. **Marcos Vinícius Ferreira Mazoni**, portador da Carteira de Identidade nº **100.261.739-5 SSP/RS**, CPF nº **339.797.660-04**, encontrado no endereço supra, e o(a) , sediado(a) no(a) , inscrito(a) no CNPJ sob o nº , doravante denominado(a) **EMPRESA**, neste ato representado(a) na forma de seu estatuto/contrato social pelo(a) Sr(a). , celebram o presente **CONTRATO DE TRANSFERENCIA DE KNOW HOW**, decorrente do Edital de Oferta Pública nº 01/2013, sujeitando-se às normas das Leis nºs 8.666/93, 9.279/96, 10.973/04, regulamentada pelo Decreto nº 5.563/05, e 10.406/02 – Código Civil, no que couber, e às cláusulas e condições seguintes:

CLÁUSULA PRIMEIRA – DO OBJETO

Avenida Antônio Carlos, 6627 – Unidade Administrativa II - 2º andar - sala 2008/2017
CEP: 31270-901 - Belo Horizonte - MG
Fone: (31) 3409-4033/3409-6465, Telefax: 3409-6030-E-mail: info@ctit.ufmg.br
<http://www.ufmg.br/ctit>



UFMG

Constitui objeto do presente contrato a transferência, a título oneroso, sem exclusividade, pela **UFMG** e o **SERPRO** à **EMPRESA**, dos direitos para **uso, produção, comercialização, prestação de serviços**, relacionados ao *Know-How* intitulado “**Práticas de Mineração de Dados Escalável para Sistemas de Governo**”, descrito no anexo I, doravante denominado **KNOW-HOW**.

Parágrafo Primeiro – O **KNOW-HOW** é decorrente de pesquisa desenvolvida pela **UFMG** e o **SERPRO**.

Parágrafo Segundo - O presente instrumento não implica transferência da titularidade dos direitos relativos ao **KNOW-HOW**, que permanecem, para todos os fins, de propriedade da **UFMG** e do **SERPRO**.

Parágrafo Terceiro - A exploração comercial do **KNOW-HOW** será realizada sem exclusividade pela **EMPRESA**, em âmbito nacional e internacional.

Parágrafo Quarto - A **EMPRESA** será informado(a) sobre a existência de terceiros interessados na exploração comercial do **KNOW-HOW**, bem como da transferência a outros interessados.

CLÁUSULA SEGUNDA – DO PRAZO DE INÍCIO DA EXPLORAÇÃO COMERCIAL DO KNOW-HOW

A **EMPRESA** terá o prazo máximo de () anos, a contar da data de assinatura do presente Contrato, para iniciar a exploração comercial do **KNOW-HOW**.



Parágrafo Primeiro - O prazo previsto no *caput* poderá ser prorrogado, por meio de Termo Aditivo, mediante justificativa da **EMPRESA**, e concordância expressa da **UFMG** e o **SERPRO**.

Parágrafo Segundo - A **EMPRESA** deverá comunicar formalmente à **UFMG** e o **SERPRO** a data de início da exploração comercial do **KNOW-HOW**.

CLÁUSULA TERCEIRA – DAS OBRIGAÇÕES DAS PARTES

Constituem obrigações do presente Contrato:

Parágrafo Primeiro - Obrigações comuns das partes:

I- responsabilizarem-se pelo sigilo das informações relacionadas ao **KNOW-HOW**, devendo, para tanto, celebrar contratos de sigilo com seus respectivos empregados/servidores e demais envolvidos que, direta ou indiretamente, a ela tenham acesso, de forma a garantir a confidencialidade das informações. As informações relativas ao **KNOW-HOW** somente poderão ser reveladas mediante anuência formal da **UFMG** e do **SERPRO**;

I.1- excetuam-se da obrigação de sigilo as informações que:

- a) comprovadamente estiverem em domínio público ou, ainda, que estiverem contidas em patentes publicadas em qualquer país antes da assinatura do presente Contrato;
- b) comprovadamente sejam solicitadas pelo Poder Judiciário, Ministério Público ou demais autoridades competentes, em processo judicial ou administrativo;



II- comunicar à outra parte qualquer informação de seu conhecimento acerca da violação dos direitos de propriedade intelectual referentes ao **KNOW-HOW**, adotando, conjunta ou isoladamente, as providências extrajudiciais e/ou judiciais necessárias à defesa contra eventual uso não autorizado, por terceiros, do produto ou processo decorrente do **KNOW-HOW**, sendo que as respectivas despesas serão arcadas 100% (cem por cento) pela **EMPRESA**.

Parágrafo Segundo – Obrigações da EMPRESA:

I- arcar com todas as despesas necessárias para a exploração comercial do **KNOW-HOW**;

II- arcar com as despesas decorrentes da promoção de medidas judiciais ou extrajudiciais para a proteção contra ato de violação, por terceiros, dos direitos de propriedade intelectual referentes ao **KNOW-HOW**, mesmo que o ajuizamento das medidas tenha sido feito por iniciativa da **UFMG** e do **SERPRO**;

III- observar as recomendações e instruções técnicas da **UFMG** e do **SERPRO**, bem como a legislação relacionada ao **KNOW-HOW**, a fim de preservar sua qualidade industrial, assumindo, exclusivamente, as responsabilidades civil, penal e administrativa por ações de terceiros, decorrentes de vícios, defeitos, eventuais ilícitos ou danos decorrentes da não observância dos procedimentos técnicos adequados à fabricação e comercialização;

IV- dar imediata ciência à **UFMG** e ao **SERPRO** do recebimento de quaisquer autuações administrativas ou citações, bem como intimações relacionadas ao **KNOW-HOW**,



respondendo, pessoal e exclusivamente, por eventuais condenações que vierem a ser cominadas;

V- abster-se de adotar conduta comercial considerada ilegal, abusiva ou contrária aos interesses da **UFMG** e do **SERPRO**;

VI- responsabilizar-se pelos tributos e encargos exigíveis em decorrência da execução do presente Contrato;

VII- explorar o *Know-How* de forma suficiente para atender à demanda do mercado;

VIII- manter, durante toda a execução do Contrato, em compatibilidade com as obrigações ora assumidas, todas as condições exigidas para a presente contratação;

IX- observar na exploração comercial do **KNOW-HOW**, as recomendações e instruções técnicas da **UFMG** e do **SERPRO**, bem como da legislação vigente, assumindo exclusivamente a responsabilidade civil, penal e administrativa por eventuais ilícitos ou danos decorrentes da não observância dos procedimentos técnicos adequados à comercialização do *Know-How*.

X- apresentar plano de negócios e executar aprimoramentos, quando necessários, mantendo a **UFMG** e o **SERPRO** sempre informados e atualizados das evoluções.

Parágrafo Terceiro - Obrigações da UFMG e do SERPRO:

I- fornecer todas as informações, documentos e material necessários para o acesso ao **KNOW-HOW**;



II- Fornecer à **EMPRESA** suporte técnico-científico associado ao **KNOW-HOW**, sendo que as condições desse suporte, inclusive o valor da respectiva remuneração a ser paga à **UFMG** e ao **SERPRO**, serão estabelecidas em instrumento jurídico próprio;

CLÁUSULA QUARTA– DAS REMUNERAÇÕES

A **EMPRESA** pagará à **UFMG** pela transferência do **KNOW-HOW** os valores abaixo discriminados:

I- O valor de R\$ _____, a título de remuneração inicial para acesso ao **KNOW-HOW**, conforme proposta ao Edital.

II- O percentual de _____ % (_____ por cento) sobre a receita líquida auferida na comercialização do **KNOW-HOW**, a título de *royalties*, conforme proposta ao Edital.

Parágrafo Primeiro - Para os fins do inciso II, *supra*, considera-se como “receita líquida”, o valor bruto auferido com a exploração comercial do **KNOW-HOW**, deduzidos os tributos incidentes sobre a operação de venda e prestação de serviços e os valores relativos às vendas e serviços cancelados, devidamente comprovados.

Parágrafo Segundo - O pagamento da remuneração prevista no inciso II deverá ser efetuado até o _____ (_____) dia útil, após o encerramento de cada semestre, acompanhado de relatório demonstrativo que especifique a quantidade de produtos e/ou serviços comercializados nesse período, o valor bruto das vendas, a especificação e o



UFMG

valor das deduções permitidas, conforme parágrafo primeiro, e o valor líquido das vendas e prestação de serviços.

Parágrafo Terceiro - O início da contagem do semestre, previsto no Parágrafo Segundo, dar-se-á a partir da data em que o(a) formalizar à **UFMG** e ao **SERPRO** o início da exploração comercial do **KNOW-HOW**, conforme previsto no *caput* da Cláusula Segunda.

Parágrafo Quarto - O pagamento de 50% dos valores determinados na presente Cláusula deverão ser efetuados pela **EMPRESA** por meio de Guia de Recolhimento Único - GRU a ser emitida pela **UFMG** na época dos respectivos pagamentos.

Parágrafo Quinto - O pagamento de 50% dos valores determinados na presente Cláusula deverão ser efetuados pela **EMPRESA** por meio de Guia de Recolhimento Único - GRU a ser emitida pelo **SERPRO** na época dos respectivos pagamentos.

Parágrafo Sexto - Caso a **UFMG** ou **SERPRO** venham a adquirir ou contratarem serviços da **EMPRESA** relacionados ao Know-how não incidirá no preço *royalties* previstos no subitem II desta cláusula, devendo a **EMPRESA** praticar preço que desconte o percentual de % (por cento).

CLÁUSULA QUINTA- DOS JUROS E DA CORREÇÃO MONETÁRIA

O atraso no pagamento das remunerações previstas na Cláusula Quarta, pela **EMPRESA**, implicará cobrança de juros de 1% (um por cento) ao mês, sobre o valor do débito atualizado, considerado o período compreendido entre a data do vencimento e a



data do efetivo pagamento, com base na variação do IPC - FIPE ou outro índice legal que porventura venha a substituí-lo, sem prejuízo das penalidades previstas na Cláusula Treze.

CLÁUSULA SEXTA - DA FISCALIZAÇÃO E AUDITORIA

A **EMPRESA** deverá manter em sua sede registros contábeis e certidões fiscais que permitam à **UFMG** e ao **SERPRO**, a qualquer tempo, sem prévia comunicação, seja através de representantes designados para este fim ou de auditores contratados, comprovar as informações relativas à produção e a comercialização do **KNOW-HOW**, bem como todas as condições exigidas para a presente contratação.

Parágrafo Único - A **EMPRESA** deverá permitir à **UFMG** e ao **SERPRO** ou a terceiro por estes indicado, a qualquer tempo, ainda que o presente Contrato seja extinto, o exame e fiscalização do uso do processo de fabricação dos produtos obtidos do **KNOW-HOW**.

CLÁUSULA SÉTIMA – DO USO DO NOME DA UFMG E DO SERPRO

A **EMPRESA** não poderá utilizar o nome da **UFMG** e do **SERPRO**, de seus departamentos, laboratórios, funcionários, pesquisadores ou estudantes, em qualquer tipo de material promocional e/ou de propaganda, sem aprovação por escrito do seu órgão competente, devendo as condições de uso, se for o caso, ficarem estabelecidas em instrumento específico.

CLÁUSULA OITAVA – DAS INOVAÇÕES TÉCNICAS



A **EMPRESA** deverá comunicar formal e imediatamente à **UFMG** e ao **SERPRO**, toda e qualquer criação, modificação ou aperfeiçoamento que gere inovação ao **KNOW-HOW**.

Parágrafo Primeiro – Ocorrendo inovação do **KNOW-HOW**, nos termos do *caput* desta Cláusula, a **EMPRESA** não poderá, isoladamente, formular o respectivo depósito do pedido de proteção, no âmbito nacional e/ou internacional.

Parágrafo Segundo - A **UFMG**, o **SERPRO**, e a **EMPRESA** figurarão como cotitulares dos direitos de propriedade intelectual decorrentes da inovação desenvolvida somente pela **EMPRESA** e, também, aquela desenvolvida em cooperação com a **UFMG** e o **SERPRO**, nos termos a serem definidos em instrumento jurídico próprio.

Parágrafo Terceiro - Verificada a hipótese prevista no *caput*, a **UFMG**, o **SERPRO** e a **EMPRESA** comprometem-se a manter o sigilo necessário à proteção da propriedade intelectual, ficando a **EMPRESA** (**definir qual das partes**) responsável pelos procedimentos necessários à proteção da inovação do **KNOW-HOW**.

Parágrafo Quarto - A **EMPRESA** ficará responsável pelos custos de depósito, manutenção e proteção das inovações tecnológicas de que trata o *caput* em âmbito nacional e internacional, se for o caso.

Parágrafo Quinto - As criações, modificações e aperfeiçoamentos que gerem inovação do **KNOW-HOW**, nos termos previstos no *caput*, poderão ser exploradas comercialmente pela **EMPRESA**, devendo as condições de exploração, inclusive o pagamento de *royalties* à **UFMG** e ao **SERPRO**, ser definidas em instrumento jurídico próprio.



UFMG

CLÁUSULA NONA – DO SUBLICENCIAMENTO DO *KNOW-HOW*

A **EMPRESA** poderá, sob sua exclusiva responsabilidade, sublicenciar, no todo ou em parte, os direitos para ***uso, produção, comercialização, prestação de serviços*** do ***KNOW-HOW*** a terceiro interessado, respeitado o prazo de vigência do presente instrumento, desde que prévia e expressamente autorizado pela **UFMG** e o **SERPRO**.

Parágrafo Primeiro - O terceiro interessado para o qual o ***KNOW-HOW*** tenha sido sublicenciado deverá respeitar todas as cláusulas e condições do presente instrumento, o que deverá constar no contrato de sublicenciamento.

Parágrafo Segundo - A **EMPRESA** será solidariamente responsável perante a **UFMG** e o **SERPRO** pelo cumprimento das cláusulas e condições estabelecidas no presente Contrato, por parte do sublicenciado, inclusive aquelas referentes à remuneração e ao sigilo.

Parágrafo Terceiro - A **EMPRESA** se compromete a enviar à **UFMG** e ao **SERPRO**, imediatamente após sua assinatura, uma via do(s) contrato(s) de sublicenciamento, bem como de seus eventuais aditamentos, que deverão, também, ter a prévia anuência da **UFMG**.

Parágrafo Quarto - Caberá à **UFMG** e ao **SERPRO** % (por cento) para cada um, sobre o valor a ser pago pelo sublicenciamento do ***KNOW-HOW*** e das inovações técnicas, sem prejuízo do recebimento pela **UFMG** e o **SERPRO** dos *royalties* previstos no inciso II, da Cláusula Quinta.

CLÁUSULA DEZ – DA TERCEIRIZAÇÃO

Avenida Antônio Carlos, 6627 – Unidade Administrativa II - 2º andar - sala 2008/2017

CEP: 31270-901 - Belo Horizonte - MG

Fone: (31) 3409-4033/3409-6465, Telefax: 3409-6030-E-mail: info@ctit.ufmg.br

<http://www.ufmg.br/ctit>



UFMG

Caso a **EMPRESA** necessite terceirizar quaisquer de suas atividades relativas à **TECNOLOGIA**, deverá comunicar formalmente à **UFMG** e ao **SERPRO** o nome da empresa ou profissional terceirizado, no prazo máximo de () dias (fixar) contados da terceirização.

Parágrafo Primeiro - A **EMPRESA** será responsável perante a **UFMG** e o **SERPRO** pelas atividades realizadas pelo terceirizado.

Parágrafo Segundo - O terceirizado não poderá, sob qualquer hipótese, utilizar o **KNOW-HOW** para outro fim, senão para o exercício das atividades para o qual foi contratado. A **EMPRESA** será responsável pela atuação e fiscalização do terceirizado, respondendo, inclusive, pela prática de atos que infrinjam o disposto no presente Contrato.

Parágrafo Terceiro – A **EMPRESA** deverá firmar termo de sigilo com o terceirizado, de forma a obrigá-lo a manter absoluto sigilo sobre toda e qualquer informação relativa ao **KNOW-HOW** e/ou suas inovações a que tiver acesso no exercício de suas funções, não podendo promover qualquer tipo de divulgação, seja a que título for, sem que haja prévia e expressa autorização da **UFMG** e o **SERPRO**.

Parágrafo Quarto - A terceirização das atividades não prejudicará o recebimento integral dos valores devidos pela **EMPRESA** à **UFMG** e o **SERPRO**.

CLÁUSULA ONZE - DA EXTINÇÃO DO CONTRATO

Constituem hipóteses de extinção do presente Contrato:



UFMG

I- rescisão, que poderá ocorrer, a critério da parte inocente, caso haja descumprimento de quaisquer de suas cláusulas ou condições;

II- resolução, em virtude de caso fortuito ou força maior, devidamente comprovados, bem como no caso em que a **EMPRESA** verificar a inviabilidade da produção ou comercialização do **KNOW-HOW**, o que deverá constar devidamente fundamentado em relatório técnico a ser avaliado pela **UFMG** e o **SERPRO**;

III- resilição, por livre acordo das partes, por meio de distrato, no qual serão estabelecidas as condições de extinção.

Parágrafo Primeiro - Em caso de rescisão, a parte culpada deverá indenizar a parte inocente por eventuais perdas e danos e lucros cessantes.

Parágrafo Segundo - A decretação de falência da **EMPRESA** constitui motivo para rescisão contratual, sem prejuízo do cumprimento das obrigações assumidas até o momento da extinção, incluído o recebimento dos valores devidos à UFMG.

Parágrafo Terceiro - A resolução prevista no inciso II dar-se-á sem quaisquer ônus para as partes e sem a devolução dos valores pagos pela **EMPRESA** à **UFMG** e o **SERPRO**, até a data da resolução.

Parágrafo Quarto - Em quaisquer das hipóteses de extinção previstas na presente cláusula, a titularidade do **KNOW-HOW** e o recebimento dos valores porventura pendentes, especialmente os relativos aos *royalties*, estarão assegurados à **UFMG** e o **SERPRO**.



Parágrafo Quinto – Ocorrendo a extinção contratual nos termos desta cláusula, a licenciada deverá devolver todos os documentos (**desenhos, informações, certificados, especificações técnicas**) que sejam de propriedade da UFMG e do **SERPRO**, no prazo de **(fixar)** dias corridos, contados da data da extinção.

CLÁUSULA DOZE – DAS PENALIDADES

O descumprimento pela **EMPRESA** de cláusulas e/ou condições do presente Contrato ensejará a aplicação das seguintes penalidades:

I- advertência;

II- multa de 2% (dois por cento) sobre o valor devido, calculada proporcionalmente aos dias de atraso, considerando o período de 30 (trinta) dias - *pro rata die* - pelo atraso no pagamento dos valores previstos nos incisos II a IV, do Parágrafo Segundo, da Cláusula Terceira, e incisos I e II, do *caput*, da Cláusula Quinta;

III- multa de 100% (cem por cento) sobre o valor previsto no inciso I, do *caput*, da Cláusula Quinta pelo descumprimento da obrigação de sigilo;

IV- multa de 50 % (cinquenta por cento) sobre o valor previsto no inciso II, da Cláusula Quinta, por descumprimento contratual não enquadrado nas hipóteses previstas nos incisos II e III desta Cláusula;

V- suspensão temporária de participação em licitação e impedimento de contratar com a Administração Pública Federal, por prazo não superior a 2 (dois) anos;



VI- declaração de inidoneidade para licitar ou contratar com a Administração Pública, enquanto perdurarem os motivos determinantes de punição ou até que seja promovida a reabilitação perante a **UFMG** e o **SERPRO**.

Parágrafo Primeiro - A aplicação das penalidades previstas nesta Cláusula poderá ocorrer sem prejuízo do disposto na Cláusula Onze.

Parágrafo Segundo - A aplicação de uma das penalidades estabelecidas nos incisos desta Cláusula não exclui a possibilidade de aplicação de outras.

Parágrafo Terceiro - Previamente à aplicação das penalidades, a Contratada será notificada por escrito, garantindo-lhe ampla defesa.

Parágrafo Quarto – As sanções previstas nos incisos V e VI poderão ser aplicadas, também, nas hipóteses previstas no art. 88, da Lei nº 8.666/93.

CLÁUSULA TREZE - DAS DISPOSIÇÕES GERAIS

I- Este Contrato obriga as partes e seus eventuais sucessores a qualquer título.

II - Qualquer aceitação, prorrogação ou tolerância da **UFMG** e do **SERPRO**, em relação às obrigações assumidas pela **EMPRESA** no presente Contrato, não constituirá alteração ou novação contratual.



III - As alterações deste instrumento que porventura se fizerem necessárias, com exceção de seu objeto, serão formalizadas tão-somente por meio de termo aditivo.

IV - O transferência do **KNOW-HOW** objeto do presente Contrato não constitui impedimento para que a **UFMG** e o **SERPRO** continuem a realizar o desenvolvimento de pesquisas relacionadas ao **KNOW-HOW**.

V- Ocorrendo fusão, cisão ou incorporação, bem como outras formas de alteração social, mudança de finalidade ou estrutura da **EMPRESA**, a **UFMG** e o **SERPRO** deverão ser comunicadas, de imediato e formalmente, acerca de tais ocorrências. Após tal comunicação, a **UFMG** e o **SERPRO** procederão, em face dos interesses da Administração, à avaliação da possibilidade de continuidade da execução do Contrato, devendo manifestar-se, com a devida motivação, pela manutenção do Contrato ou pela sua rescisão.

VI – A cessão total ou parcial do Contrato deverá ser formalmente solicitada à **UFMG** e ao **SERPRO**. Após o devido exame, sopesados os interesses da Administração e comprovado que a continuidade da execução do Contrato poderá ocorrer nos mesmos termos e condições pactuados e que não lhe acarretará prejuízos, a **UFMG** e o **SERPRO** decidirão, motivadamente, pelo deferimento do pleito. Em caso negativo, com a devida justificativa, a **UFMG** e o **SERPRO** decidirão pela rescisão contratual.

VII- Qualquer notificação acerca da execução deste Contrato, a ser feita pelas partes envolvidas umas às outras, poderá ser entregue pessoalmente ou enviada por meio de FAX ou e-mail, cujo original, devidamente assinado, deverá ser postado até o dia



seguinte, pelo correio, com aviso de recebimento, no endereço respectivo da parte notificada, conforme se segue:

- a **UFMG**: Avenida Antônio Carlos, 6627, Unidade Administrativa II, 2º andar, sala 2013, Belo Horizonte, Minas Gerais;

- o **SERPRO**: SGAN-Quadra 601, Módulo “V”, CEP: 70836-900, Asa Norte, Brasília, Distrito Federal ;

- a **EMPRESA** . : .

CLÁUSULA QUATORZE - DA VIGÊNCIA

O presente instrumento terá vigência de 5 (cinco) anos, a contar da data de sua assinatura, podendo ser prorrogado pelas partes mediante assinatura de termo aditivo.

CLÁUSULA QUINZE - DA PUBLICIDADE

Caberá a **UFMG** proceder à publicação do extrato do presente Contrato na Imprensa Oficial, no prazo estabelecido no Parágrafo Único, do art. 61, da Lei 8.666/93.

CLÁUSULA DEZESSEIS - DO FORO

O foro competente para dirimir dúvidas ou litígios oriundos deste instrumento é o da Justiça Federal, Seção Judiciária de Minas Gerais, nos termos do inciso I, do art. 109, da Constituição Federal.



UFMG

E, assim, por estarem justas e acordadas, firmam o presente, em 3 (três) vias, de igual teor e forma, para os mesmos efeitos legais, na presença das testemunhas a seguir assinadas.

Belo Horizonte, ____ de _____ de 20____.

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Pró-Reitor de Pesquisa

SERVIÇO FEDERAL DE PROCESSAMENTO DE DADOS

Marcos Vinícius Ferreira Mazoni

EMPRESA

TESTEMUNHAS:

1- _____

2- _____

Nome:

Nome:

CPF:

CPF:

Avenida Antônio Carlos, 6627 – Unidade Administrativa II - 2º andar - sala 2008/2017

CEP: 31270-901 - Belo Horizonte - MG

Fone: (31) 3409-4033/3409-6465, Telefax: 3409-6030-E-mail: info@ctit.ufmg.br

<http://www.ufmg.br/ctit>