

Prospecção sobre a Extração da Personalidade do Usuário e seu uso em Sistemas Computacionais

G. G. Cardoso¹; I. da C. Carvalho¹; M. A. S. N. Nunes²;

¹Universidade Federal de Sergipe – DCOMP (UFS), São Cristóvão – Sergipe, Brasil.

gcardoso18@gmail.com

¹Universidade Federal de Sergipe – DCOMP (UFS), São Cristóvão – Sergipe, Brasil.

iraildodacosta@gmail.com

²Universidade Federal de Sergipe – DCOMP/PROCC (UFS) e PPGPI, São Cristóvão – Sergipe, Brasil.

gutanunes@gmail.com

(Recebido em 12 de fevereiro de 2014; aceito em 01 de julho de 2014)

A personalidade do ser humano ainda é objeto de muitos estudos na psicologia, com diversas teorias e sem algum consenso. Mas uma certeza é que a personalidade revela diversas informações sobre uma pessoa. Com esse intuito, existem pesquisas em como estruturar a personalidade do usuário num formato computacional a fim de usá-la num sistema. Pois seu uso poderia fazer com que diversos sistemas, como o de recomendações ou interfaces fosse mais adequada ao gosto ou às necessidades do usuário. O objetivo desta prospecção é buscar indícios sobre o estado atual da técnica via o cenário de patentes depositadas relacionadas a produtos que utilizem a personalidade e a extração da mesma no contexto computacional.

Palavras-chave: personalidade, extração de personalidade, personalidade em sistemas computacionais.

Technological forecasting about the use and extraction of User's Personality in Computing Systems

The personality of human still is object of many studies in the psychology, based on many theories but without consensus among researchers. However a certainty is that the personality reveals much different information about a person. Towards this, there are many researches going on about how to organize the user's personality in a computational format to be used in a system. Because its use could be more adequate to many systems such as recommender systems or software that personalize interfaces being more adequate to the users' needs. The objective of this technological forecasting is verify the current state of the technique about patents' scenario in order to wonder about new products that uses personality and its extraction on the computational context.

Keywords : personality, personality's extraction, personality in computing systems.

1. INTRODUÇÃO

A personalidade humana é objeto de vários estudos como o de Goldberg¹, que propõe um modelo para representar a personalidade. Tal modelo é conhecido como *Big Five*, que modela a personalidade baseado em cinco fatores, sendo eles: Consciência, Agradabilidade, Neuroticismo (ou Estabilidade Emocional), Abertura à novas Experiências e Extroversão. Todos os fatores juntos representam as dimensões da personalidade do indivíduo, segundo Goldberg¹. A abordagem computacional do uso de personalidade foca na teoria da personalidade com base na ênfase estrutural, considerando que existem quatro grupos teóricos e suas ênfases (psicodinâmica, realidade percebida, aprendizagem e estrutural), segundo trabalho de Goldberg¹. Como as teorias com ênfase estrutural possuem uma preocupação maior em como se estrutura a personalidade, isso as torna mais interessantes para serem estudadas no ramo da computação. A partir dessas teorias, também foi possível construir métodos para extrair a personalidade de um indivíduo.

Objetivando a extração de personalidade de indivíduos, testes de mensuração de personalidade se tornam úteis. Os mesmos podem atribuir valores a características da personalidade de uma pessoa. A personalidade depois de extraída pode ser utilizada de várias formas: para decidir o indivíduo mais adequado para um determinado cargo com base no perfil psicológico, lojas podem oferecer produtos e

serviços que melhor se adequem ao usuário, ou reunir pessoas em grupos por afinidades (seja profissionalmente, socialmente), por exemplo, como pode ser visto nos trabalhos de Nunes^{2,3}.

Segundo Gosling⁴, as pessoas deixam marcas de sua personalidade em seus espaços pessoais, tal como quartos, escritórios, etc. E com a ascensão de redes sociais, existe a possibilidade de que vários traços da personalidade do usuário estejam no perfil do mesmo como pistas nessa redes. Assim torna-se promissor procurar métodos de extrair a personalidade ou parte dela a partir das informações contidas nesses perfis.

Pesquisas no sentido de extrair a personalidade em mídias sociais estão sendo realizadas por Quercia *et al.*⁵, por exemplo. Eles apresentam em seus trabalhos, a possibilidade de extrair parte da personalidade do usuário com base das informações públicas contidas no perfil dele no *Facebook*®, as informações públicas utilizadas para extração são: quantidade de amigos, páginas curtidas, quantidade de postagens e marcações que recebe e faz. Por exemplo, outro trabalho de Quercia *et al.*⁵ é sobre extrair a personalidade com base no perfil do *Twitter*®. Quercia *et al.* usam as seguintes informações: quantidade de usuários seguidos pelo usuário, quantos o seguem, quantidade de *tweets* realizados pelo usuário. Eles ainda mostram que existem algumas correlações dessas informações com alguns fatores da personalidade no modelo *Big Five*. Quercia *et al.* vem mostrando que é possível encontrar novas formas de extrair a personalidade utilizando meios menos custosos ao usuário.

Considerando que extração de personalidade é uma área em crescimento e com potencial mercadológico esse artigo tem como finalidade realizar uma prospecção para tomar conhecimento das perspectivas existentes no estado da técnica atual com relação a extração e uso da personalidade no âmbito computacional. Uma prospecção é um estudo de registros de patentes para analisar o mercado atual com relação à tecnologia aplicada e verificar se o mesmo já está saturado para um lançamento de algum futuro produto (RUSSO *et al.*, 2012)⁶.

O artigo está estruturado como segue: na seção 2 é apresentada a metodologia usada no artigo; na seção 3, apresenta-se a análise dos dados; na seção 4, conclui-se o trabalho; finalizando pelas referências.

2. METODOLOGIA

A prospecção foi realizada na base internacional de patentes *Derwent Innovation Index*⁷ no mês de julho de 2013. A base *Derwent* possui registros de depósito de patentes de vários órgãos emissores de patentes, tais como dos escritórios americanos, europeus, alemães e japoneses *Derwent-Ajuda*⁸. Foram usadas as seguintes palavras-chave nessa prospecção: *Personality Mining*, *Personality Data Structure*, *Personality Mobile Measurement*, *Personality Stories*, *Personality Research* e *Personality Traits*. Os campos utilizados na base *Derwent* foram o de título (*title*) e resumo/tópico (*topic*). Apenas resultados que fossem filtrados para o campo de Engenharias e Computação foram considerados.

A motivação para o uso do *Derwent* foi devido ao fato do *mesmo* possuir links diretos para a busca de textos completos, incluindo, também, diversas facilidades para gravação de registros e disponibilização dos dados considerando também o contrato oficial com 40 organismos internacionais e nacionais de registro e concessão de patentes, segundo Nunes⁹.

3. ANÁLISE DE DADOS E DISCUSSÃO

Realizada a busca por patentes, notou-se um retorno maior de depósitos de patentes por resumo e por título, relacionados com os termos *Personality Traits* (sessenta e cinco patentes), seguindo por *Personality Research* (trinta e quatro patentes), e *Personality Data Structure* (vinte e cinco patentes). Outros termos utilizados (*Personality Mining*, *Personality Mobile Measurement* e *Personality Stories*) retornaram resultados pouco expressivos comparados aos três primeiros. Vale frisar que a quantidade de registros retornados foi significativamente superior que aos apresentados aqui, mas não pertenciam às categorias da Engenharia ou Computação, fazendo menção a medicamentos ou registros na área da Psicologia; logo esses resultados apresentados na Tabela 1 são filtrados apenas para as áreas de

Engenharia ou Computação. A Tabela 1 apresenta os resultados da quantidade de patentes depositadas relativas aos termos pesquisados:

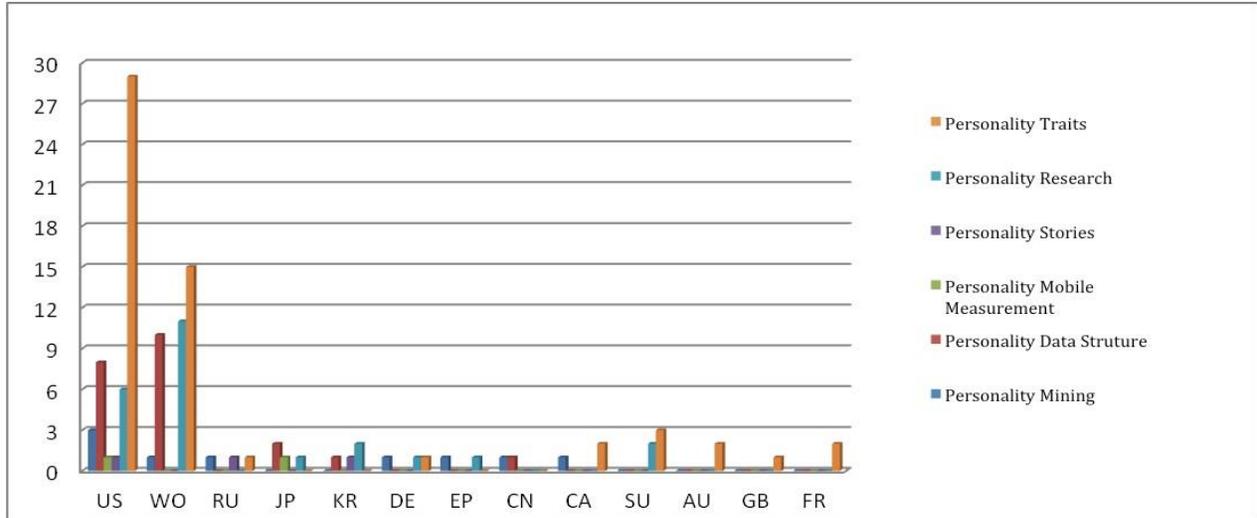
Tabela 1: Quantidade de registros obtidos com os termos, por tópico e título.

Termo	Resumo	Título	Total (Resumo + Título)
Personality Mining	5	1	6
Personality Data Structure	34	0	34
Personality Mobile Measurement	2	0	2
Personality Stories	2	0	2
Personality Research	22	3	25
Personality Traits	51	14	65

Fonte: Autoria própria (2013).

Sobre a quantidade de patentes, a Figura 1 mostra a quantidade de patentes depositadas por países com os termos pesquisados nos resumos, onde o total equivale à soma de patentes depositadas encontradas em resumos e títulos para o termo pesquisado; os Estados Unidos foi a nação com maior quantidade de patentes depositadas, seguindo apenas pelo conjunto de patentes depositadas via PCT. Excluindo-se essas, os países que mais depositaram patentes foram a antiga União Soviética com cinco patentes (alguns dos registros mais antigos foram registrados enquanto ela ainda existia), Japão e Coreia com quatro patentes cada. É possível inferir com estes resultados que há um grande enfoque de produtos criados (que usam personalidade) principalmente nos EUA, considerando as palavras-chave acima.

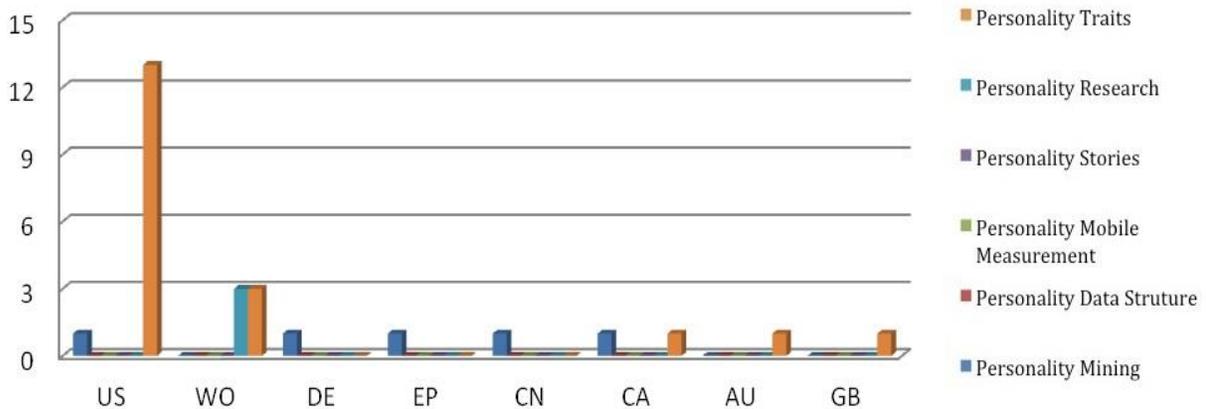
A legenda de cores para as Figuras é apresentado abaixo:



Significado das siglas: US= Estados Unidos, WO= depósitos via PCT, RU= Rússia, JP= Japão, KR= Coreia do Sul, DE= Alemanha, EP= depositado via Organização Europeia de Patentes, CN= China, CA= Canadá, SU= União Soviética, AU= Austrália, GB= Reino Unido e FR= França.

Figura 1: Quantidade de Patentes por País(Tópico).
Fonte: Autoria própria(2013).

A Figura 2 apresenta a quantidade de patentes depositadas por país de origem baseado nas palavras chave pesquisadas nos títulos, ficando evidente a maior quantidade de registros depositados pelos EUA. Tal fato também é devido à maioria das patentes encontradas pertencentes aos EUA serem provenientes da palavra-chave *Personality Traits*. Poucas patentes foram retornadas se comparadas à pesquisa por resumo, somente as palavras-chave *Personality Mining*, *Personality Traits* e *Personality Research* retornaram patentes.



Significado das siglas: US= Estados Unidos, WO= depósitos via PCT, DE= Alemanha, EP= depósitos via Organização Europeia de Patentes, CN= China, CA= Canadá, AU= Austrália e GB= Reino Unido.

Figura 2: Quantidade de Patentes por País(Título).
Fonte: Autoria própria(2013).

Analisando a quantidade de depósitos de patentes por ano, no geral foi percebido um aumento no depósitos de patentes com foco em todas as palavras chave que foram usadas na prospecção. Valendo frisar o ano de 2008 como sendo o ano em que houve a maior quantidade de depósitos (catorze no total),

sendo seguido por uma queda de registros (apenas três registros no ano de 2009), nos anos seguintes houve um novo crescimento na quantidade de depósitos (2010 a 2012). Na última década a quantidade de patentes depositadas foi notadamente maior do que a quantidade de patentes depositadas antes do ano 2000. Houve oitenta e seis patentes depositadas entre os anos de 2001 e 2012, e catorze registradas antes disso, mostrando um grande interesse recente na área. O termo *Personality Traits* é o que mostra maior concentração de depósitos. Os termos *Personality Stories* e *Personality Mining* foram pouco expressivos em quantidades de depósitos, apesar de existirem depósitos do ano de 2001, a maior parte dos registros retornados pelas palavras chave é do ano 2008 em diante, como apresentado na Figura 3.

Na Figura 3 são apresentadas as quantidades de depósitos por ano, vale informar que algumas datas não aparecem no gráfico, pois não houve depósitos, sendo assim são apresentados somente os anos em que houveram depósitos.

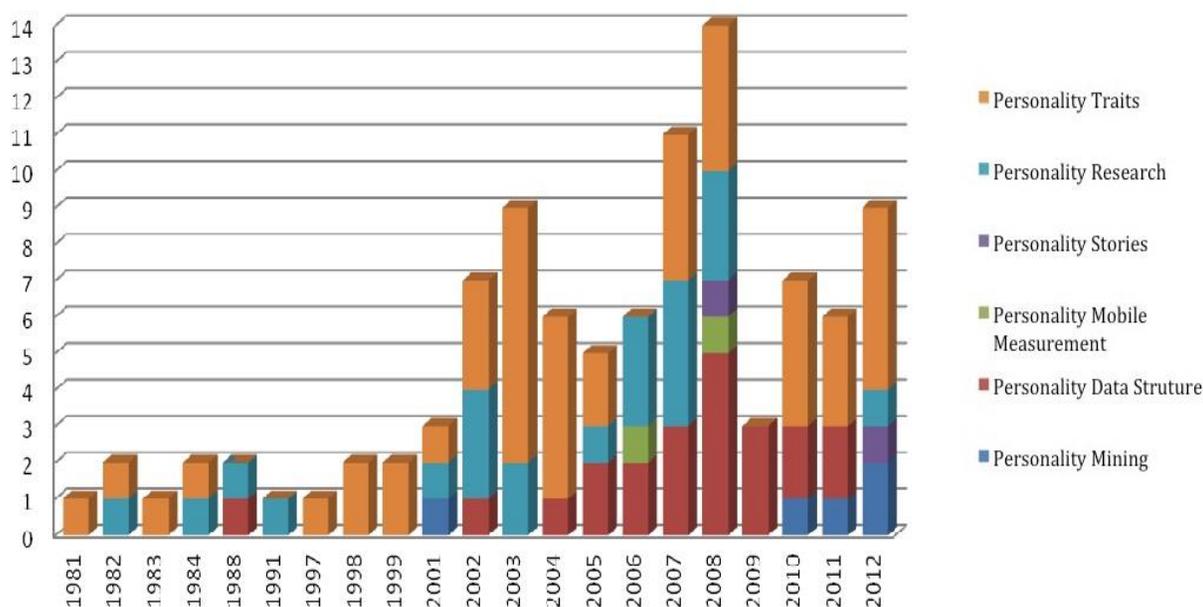


Figura 3: Quantidade de Patentes por ano e palavras chave.

Fonte: Autoria própria(2013).

Na Figura 4, sobre a quantidade de depósitos pelos CIPs*, como não há uma seção específica para Computação ou Engenharia, muito provavelmente estão agrupadas na seção de Física, o que explica a imensa quantidade na Seção G (referente à Física), pois só um CIP (o G06F) possui mais de 70 depósitos. A classe de Cálculo/Contagem (classe 06) apresentou mais depósitos, pois as três que mais possuem registros além de serem da área da Física, são desta classe, sendo as três subclasses em ordem decrescente de quantidade de registros: Processamento Elétrico de Dados Digitais, Identificação de Dados e Métodos de Processamento de Dados. Também foram encontrados depósitos de áreas não esperadas, como Necessidades Humanas (Classe A), Construções Fixas (Classe E), Química e Metalurgia (Classe C). As classes com maior número de depósitos foram a G06F que comporta “processamento elétrico de dados digitais; a G06K, que comporta “identificação de dados; apresentação de dados; suporte de dados; manipulação de transporte de dados”; a G06 Q, que comporta “sistemas ou métodos de processamento de dados, especialmente adaptados para propósitos administrativos, comerciais, financeiros, de

*O CIP, ou Código Internacional de Patente, é um código não único válido mundialmente que é atribuído a uma patente para que esta seja identificada e enquadrada nas categorias existentes; primordialmente é constituído de quatro caracteres, sendo o primeiro uma letra representando a seção da patente, vindo logo após mais dois caracteres que são representados por números, onde o conjunto de três caracteres forma a classe da patente; e por fim, o último caractere é uma letra representa a subclasse⁶.

gerenciamento, supervisão ou predição; sistemas ou métodos especialmente adaptados para propósitos administrativos, comerciais, financeiros, de gerenciamento, supervisão ou predição, não incluídos em outro local”, e; a G09B, que comporta “aparelhos educativos ou de demonstração; aparelhos para ensino ou comunicação com os cegos, surdos ou mudos; modelos; planetários; globos; mapas; diagramas”.

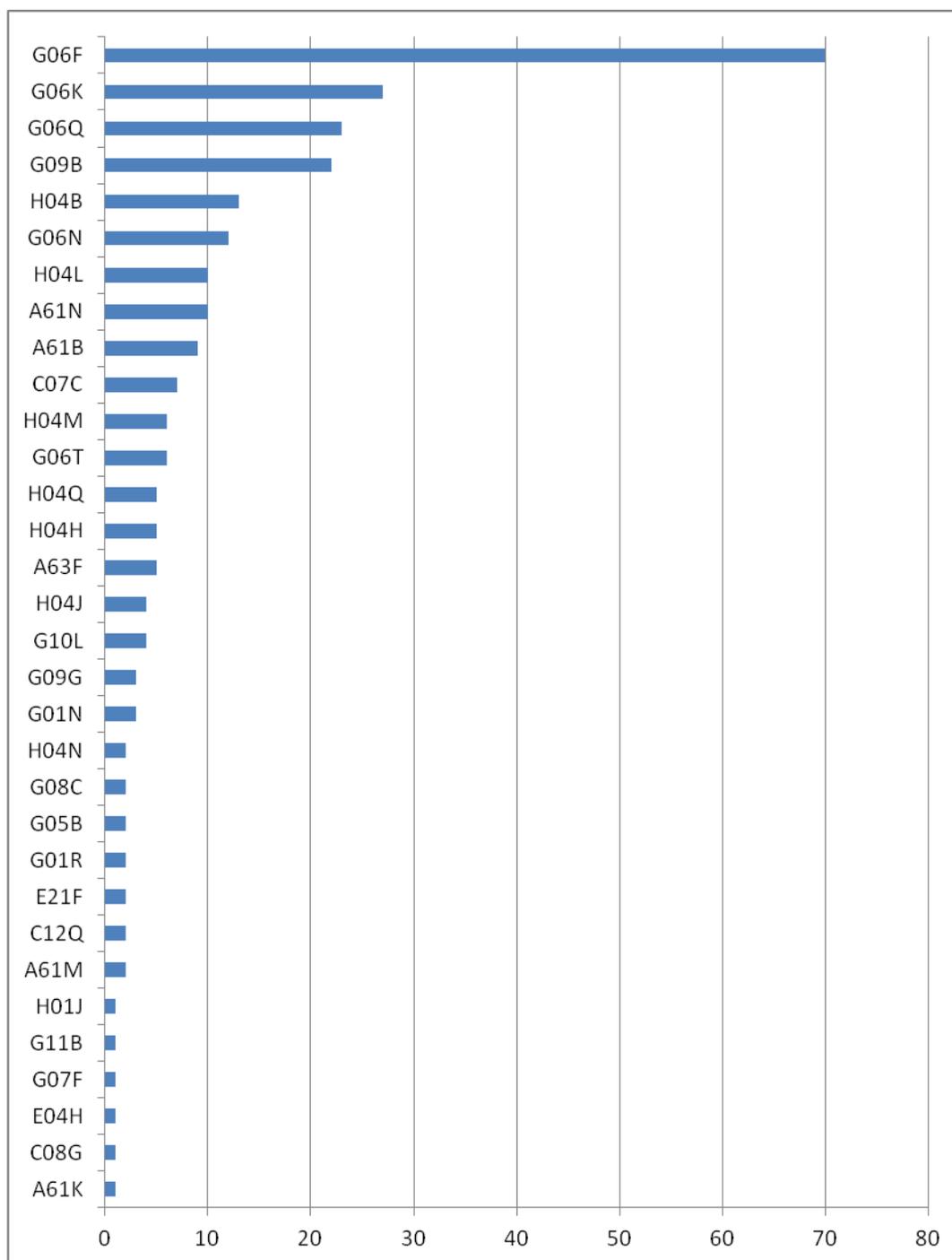


Figura 4: Quantidade de Patentes por CIP

Fonte: Autoria própria(2013).

4. CONCLUSÃO

A prospecção mostrou que apesar de um aumento significativo no depósito de patentes nos últimos anos, a quantidade de depósitos relevantes ainda é muito baixa, não há muitas patentes na linha de extração e uso da personalidade do indivíduo em sistemas computacionais. Mesmo que a busca de patentes depositadas tenha sido refinada para as áreas de Engenharia e Computação, foram retornados depósitos da área de necessidades humanas (seção A), por exemplo, que não condiziam com a demanda da prospecção; podendo ser em função da escolha das palavras chave.

Percebe-se também que a maior parte dos depósitos de patentes fez referência à seção G do CIP, que significa a seção de produtos relacionados à Física; como não há uma seção específica para Computação, provavelmente essa é a que mais se aproxima (foi verificado que os depósitos de dispositivos computacionais pertenciam a esta seção). Outro detalhe é que a classe 06 (sobre cálculos e contagem de dados) continha a maior parte dos depósitos advindos da seção G.

É interessante realçar que utilizar a personalidade do usuário é uma área promissora, principalmente considerando as pistas públicas deixadas pelos indivíduos em redes sociais. Acredita-se que há um crescente potencial industrial nessa área. A prospecção apresentada nesse artigo apresenta uma tendência à esse crescimento. Entretanto os autores têm ciência que as palavras-chaves utilizadas podem não representar a realidade, considerando que no meio industrial muitas vezes não utiliza as mesmas palavras chaves usadas em buscas no estado-da-arte, isso acontece para tentar burlar a busca e replicação de tecnologia não protegida territorialmente. Os autores também tem ciência do limite da representatividade das conclusões do artigo, haja vista que a base de prospecção usada incluiu apenas o *Derwent*. Entretanto como já justificado, o *Derwent* inclui registros de depósitos de patentes proveniente de escritórios importantes do mundo, tais como escritório europeu (EPO), americano (USPTO), alemão, japonês. Outra informação importante, a base brasileira INPI não foi utilizada pois a legislação brasileira não permite depósitos de patentes de *software* ou similares, o que pode justificar a ausência de resultados vindos do Brasil, em contra-partida, patentes de *software* são permitidas na lei americana. Entretanto todo *software* com potencial internacional é patenteado nos EUA, para garantir o direito territorial, enquanto em países onde é válida a lei do direito autoral, o mesmo é protegido desde sua criação, porém sua funcionalidade não é exclusiva do autor.

-
1. Goldberg, L.R. An alternative “Description of personality”: The Big-Five factor structure. **Journal of Personality and Social Psychology**, 1990, 59, 1216-1229.
 2. Nunes, M.A.S.N. Computação Afetiva personalizando interfaces, interações e recomendações de produtos, serviços e pessoas em ambientes computacionais. In: Nunes, M.A.S.N.; Oliveira, A.A.; Ordonez, E.D.M. **Pesquisas e Projetos**. São Cristóvão: Editora UFS, 2012. 115-151.
 3. Nunes, M.A.S.N. Recommender Systems based on Personality Traits. 2008. 142f. Tese (Doutorado em Computação) – Université Montpellier 2, Montpellier, Hérault, 2008.
 4. Gosling, S. Snoop: What Your Stuff Says About You. Nova Iorque: Basic Books, 2008. 263p.
 5. Quercia, D.; Lambiotte, R.; Stillwell, D.; Kosinski, M.; Crowcroft, J. The Personality of Popular Facebook Users. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. 2012. Seattle. Anais... New York: ACM, 2012. 955-964.
 6. Russo, S.L. et al. Propriedade Intelectual. In: Russo, S. L. et al. (Org). Capacitação em Inovação Tecnológica para Empresários. São Cristóvão: Editora UFS, 2012. 55-90.
 7. Derwent Innovation Index. WEB OF KNOWLEDGE. Disponível em: <<http://isiknowledge.com/diidw>>. Acesso em: 15 de maio de 2013.
 8. Derwent Ajuda. Disponível em http://images.webofknowledge.com/WOKRS511B5/help/pt_BR/DII/hs_citation_country_coverage.html Acesso em: 15 de junho 2014.

9. Nunes, M.A.S.N. Produção Tecnológica na IE: prospecção e propriedade intelectual em Informática na Educação. In: II Jornada de Atualização em Informática na Educação (JAIE 2013). 2013. Campinas - SP. Anais... Campinas: Sociedade Brasileira de Computação, 2013. 30p.