

Implicações da categorização e indexação na recuperação da informação tecnológica contida em documentos de patentes

Anna Haydée Lanzillotti Jannuzzi

Mestre em tecnologia pelo Centro Federal de Educação Tecnológica Celso Suckow da Fonseca.
E-mail: ahaydee@int.gov.br

Rita de Cássia Rocha Amorim

Mestranda em tecnologia do Cefet/RJ.
E-mail: ritacra@inpi.gov.br

Cristina Gomes de Souza

Doutora em engenharia de produção pela Universidade Federal do Rio de Janeiro.
E-mail: cgsouza@cefet-rj.br

Resumo

A Classificação Internacional de Patentes, criada para uniformizar a sistematização dos documentos de patentes, ainda não atingiu seus objetivos. Documentos iguais são indexados de forma distinta em cada país signatário do Acordo de Estrasburgo. O objetivo do trabalho é apresentar as causas da subjetividade nas etapas envolvidas na construção da identidade das patentes de invenção, principalmente no que tange à categorização e indexação, etapas importantes para que os usuários do sistema de proteção patentária possam recuperar a informação tecnológica contida nestes ativos de propriedade intelectual. O trabalho baseou-se em revisão bibliográfica e pesquisa documental.

Palavras-chave

Patentes. Informação tecnológica. Ciência da informação. Ciência cognitiva.

Implications of the categorization and indexation in the technological information retrieval in documents of patents

Abstract

The International Classification of Patents, created for conforming the systemization of patent documents, has not yet reached its objective. The same documents are indexed in different ways by each signatory country of Strasbourg's Agreement. The objective of this paper is to present the subjective causes involved mainly in the processes of categorization and indexation that are important stages to allow that the users of a patent system can retrieve the technological information contained in these intellectual property assets. The research was based on bibliographic review and documental search.

Keywords

Patents. Technological information. Information science. Cognitive science.

INTRODUÇÃO

A chamada Era da Informação enfatizou a relação homem-máquina. Diante da complexidade do tema, emergiram novas ciências com caráter interdisciplinar que procuram compreender os fenômenos da informação e do conhecimento como a ciência cognitiva (CC) e a ciência da informação (CI). Tais ciências preocupam-se em estudar a formação e a representação do conhecimento, assim como sua assimilação e transmissão, principalmente utilizando-se das novas tecnologias de informação e comunicação (TICs), visando a adequar-se aos requisitos da sociedade globalizada.

Ao estudarem elementos comuns, CC e CI formaram uma interseção na qual são identificados quatro processos: categorização; indexação; recuperação de informação; relação homem-máquina (computador). Esses processos são empregados pelo sistema de proteção patentária. Os documentos de patentes são categorizados e indexados com o auxílio da Classificação Internacional de Patentes. Após a indexação, os documentos são inseridos em bancos de dados informatizados para que os usuários do sistema de patentes possam recuperar a informação tecnológica contida nesses ativos de propriedade intelectual. Essas etapas, por envolverem, de um lado, várias identidades guarnecidas da natureza humana e, de outro, o caráter extremamente racional das máquinas, geram questões que alteram a transmissão e a recuperação de informação, dificultando a absorção do conhecimento.

O objetivo do trabalho é apresentar as causas da subjetividade nas etapas envolvidas na construção da identidade das patentes de invenção, principalmente no que tange à categorização e indexação, etapas importantes para que os usuários do sistema de proteção patentária possam recuperar a informação tecnológica contida nestes ativos de propriedade intelectual.

A metodologia da pesquisa consistiu em revisão bibliográfica e documental, complementada com pesquisa de campo baseada em observação participante, uma vez que um dos autores apresenta vivência profissional em análise de patentes. Também foram acessados bancos de dados patentários do Brasil e do exterior.

CIÊNCIA COGNITIVA E CIÊNCIA DA INFORMAÇÃO: BASES PARA REPRESENTAÇÃO DO CONHECIMENTO

Segundo Gardner (1996, p.19), a ciência cognitiva é

um esforço contemporâneo, com fundamentação empírica, para responder a questões epistemológicas de longa data – principalmente aquelas relativas à natureza do conhecimento, seu desenvolvimento e seu emprego.

Atualmente, o objeto de estudo da ciência cognitiva se concentra nas atividades mentais, tais como o pensamento, a imaginação, a recordação, a solução de problemas, a percepção, o julgamento e a aprendizagem da linguagem, entre outras, que ocorrem diferentemente em cada indivíduo.

Estudar o conhecimento é um dos desígnios da ciência cognitiva (MEY, 1982, *apud* LIMA, 2003), assim como sua representação e suas formas de manipulação. Quando se fala em conhecimento, deve-se reportar à sua representação como um conjunto de convenções sintáticas e semânticas que viabilizam descrever os objetos. Relativamente à semântica, cabe pontuar a ocorrência da influência contextual, ou seja, a influência do mundo exterior que cada indivíduo integra e que cria uma consciência coletiva unindo seus participantes, chamada por Ballone (2003) de “concordância cultural”. Para o autor, é por meio dos processos de percepção, aprendizagem, memória e pensamento produtivo que surgem os conceitos que se configuram como um dos componentes de maior relevância para a construção do conhecimento e que são um atributo intelectual, assim como juízos e pensamento racional. Outro requisito de extrema importância na formação do conhecimento é a “inferência”, definida como a capacidade de se fazerem analogias ou comparações, diferentemente do conceito que se baseia em classificação e associação de idéias.

Dentro da ciência cognitiva, existem duas vertentes: o cognitivismo e o conexionismo. A primeira, presente também na ciência da informação, assimila o humano à máquina. Assim, a mente equivale a um computador, a uma máquina 0, e como tal, possui um repertório de instrumentos, um mapa de representações do mundo e uma linguagem que lhe permite resolver problemas. A cognição é vista como a solução. Por meio de absorção de novas experiências, aprende-se criando representações do mundo independente e externo. A corrente cognitivista importa-se em “como” a informação é processada pelo observador, considerando as informações preexistentes na mente, e em “como” essas serão arquivadas transformando-se em memória (MAGRO,

1999, *apud* BORGES *et al.*, 2003). O conexionismo, como o cognitivismo, trabalha também com a idéia de que a mente é processadora de informações. A representação não é mais inata, mas fruto das experiências obtidas no meio e do processamento da informação.

Já o nascimento da ciência da informação foi uma imposição da necessidade da difusão de informação, considerando seu aumento de volume e diversificação (CARDOSO, 1996, *apud* ARAÚJO, 2003). Foi definida, em reunião do Georgia Institute of Technology, como

a ciência que investiga as propriedades e o comportamento da informação, as forças que governam o fluxo da informação e os meios de processamento da informação para acessibilidade e usabilidade ótimas. Os processos incluem a geração, disseminação, coleta, organização, armazenamento, recuperação, interpretação e uso da informação. A área é derivada de ou relacionada à matemática, lógica, lingüística, psicologia, tecnologia computacional, pesquisa operacional, artes gráficas, comunicações, biblioteconomia, administração e algumas outras áreas (SHERA; CLEVELAND, 1977, *apud* BRAGA, 1995).

A conexão entre a ciência cognitiva e a ciência da informação reside no interesse de como a informação produz conhecimento e como é processada. A informação passa a ser reconhecida como elo entre o conhecimento e o desenvolvimento humano, uma vez que o conhecimento ocorre quando a informação é percebida e aceita, sendo considerada como a alteração provocada no estoque mental de saber do indivíduo, oriunda da interação com estruturas de informação.

De acordo com Alvarenga (2003), no escopo da ciência da informação, três estágios cognitivos encontram-se presentes correspondendo aos processos de: 1) produção, estágio anterior à entrada do documento nos sistemas de informação envolvendo a cognição dos autores, autores citados, revisores e outros; 2) organização, estágio que corresponde à inclusão do documento nos sistemas de informação, o que envolve a cognição dos profissionais responsáveis pelo tratamento da informação; 3) recuperação de conhecimento e informações, estágio pós-inclusão do documento nos sistemas de informação e que envolve a cognição dos usuários desses sistemas.

Lima (2003), por sua vez, sintetiza as áreas de interseção entre as ciências cognitiva e da informação necessárias ao processamento da informação em quatro atividades: categorização, indexação, recuperação da informação e interação homem-máquina.

A categorização ou classificação

é um processo cognitivo de dividir as experiências do mundo em grupos de entidades ou categorias, para construir uma ordem física e social do mundo” (JACOB; SHAW, 1998, p.155, *apud* LIMA, 2003, p.82).

É uma faculdade mental do indivíduo que permite interagir com o mundo e estudar os conceitos que compõem o campo de conhecimentos e as relações entre eles, estabelecendo o reconhecimento de similaridades e diferenças. É o ato de organizar o conhecimento de qualquer área que perpassa pelo estudo da representação à recuperação e, conseqüentemente, da relação entre estes conceitos. A estruturação do conhecimento facilita a recuperação da informação.

Tristão, Fachin e Alarcon (2004) acrescentam que os sistemas de classificação são linguagens documentárias que fazem a comunicação entre a linguagem natural dos usuários e a unidade de informação. Ainda segundo os autores, estudiosos de diversas áreas que têm realizado pesquisas relacionadas à representação da informação e organização do conhecimento vêm na classificação a maneira de ordenação desse conhecimento, sendo que a determinação e seleção das classes contempladas por um esquema de classificação devem estar relacionadas às necessidades de utilização daquele esquema.

Nesse sentido, cabe a distinção entre dois tipos fundamentais de classificação: as dicotômicas, baseadas no critério de presença ou ausência de determinada propriedade, e as classificações baseadas na propriedade qualificada como diferença específica. As dicotômicas tornam o ato classificatório mais fácil, por basearem-se sucessivamente em duas opções (subconjuntos) simultaneamente exclusivos. Nas classificações com base na propriedade qualificada como diferença específica, a dificuldade aumenta quanto maior for o número das várias propriedades que podem desempenhar a função de diferença específica. Assim,

a escolha de uma ou outra dessas propriedades terá como resultado diferentes arranjos das realidades a classificar e, conseqüentemente, a constituição de diferentes classificações (PERELMAN, 1963, *apud* POMBO, 2002, p.26).

Ainda há de se ressaltar que a classificação documental é uma classificação denominada por Apostel (1963, *apud* POMBO, 2002) pragmática, correspondendo a uma solução extrínseca, ou seja, que retira o foco principal do domínio a classificar, privilegiando a perspectiva do uso que o

classificador pretende dar a esse domínio e das ações que sobre ele pretende desencadear. Essa proposição é reforçada por Bhattacharyya e Ranganathan (1974, *apud* POMBO, 2002, p.30), que dizem que a diferença entre as classificações das ciências e as classificações documentais encontra-se “justamente no carácter em geral meramente especulativo das primeiras em contraste com os intuítos funcionais imediatos das segundas”, acrescentando que,

enquanto que as primeiras são esquemas globais, sistemas teóricos que não descem a detalhes nem se enredam com minúcias de classificação de domínios restritos, as segundas são propostas minuciosamente elaboradas, em geral acompanhadas de um código em que cada classe é designada por um símbolo.

Já a indexação é o processo intelectual que envolve atividades cognitivas na compreensão do texto e a composição da representação do documento. A indexação deve utilizar uma estrutura da linguagem própria para determinar a formação de categorias. O processo de indexação pode ser dividido em três etapas (LIMA, 2003):

- análise do documento e estabelecimento do seu assunto: em que observa explicitamente a base cognitiva;
- identificação dos principais conceitos do documento: em que são utilizados o intelecto do indexador e o seu conhecimento prévio armazenado na memória, que é comprometido pelas suas representações possuidoras de forte carga emotiva, influenciando as suas decisões, uma vez que percepção, emoção, atitudes, linguagem e desenvolvimento mental são considerados no estudo da formação de conceitos;
- tradução destes conceitos em termos de uma linguagem de indexação: que envolve a interpretação do conteúdo do texto.

O caráter subjetivo da indexação demonstra a complexidade de sua atividade e leva à idéia da cientificidade do trabalho desempenhado que, conforme Allen (1991, *apud* BORGES *et al.*, 2003), são atividades mentais desempenhadas de maneira distinta, por indivíduos com graus de competência diferentes em raciocínio lógico, memória visual ou experiência em vocabulário, o que pode afetar o desempenho da recuperação da informação. Lucas (1999) observa que o trabalho do indexador é promover correlação entre a terminologia do documento analisado, a terminologia do sistema de informação e a terminologia do usuário.

A recuperação da informação (RI), segundo Ingwersen (1982 *apud* LIMA, 2003, p.84) é uma tarefa que se preocupa em

trazer as estruturas cognitivas dos autores, dos *designers* de sistemas e dos indexadores de acordo com as estruturas do profissional que lida com a informação e o usuário, para cobrir a necessidade atual.

Freqüentemente, é tratada como sinônimo de ‘busca de informação’, considerando que é a necessidade da informação que promove o processo de recuperação da informação.

A RI requer que as estruturas de conhecimento do autor do texto, do indexador e a do usuário estejam em sintonia, pois são expostas três imagens da representação do mundo que necessitam serem harmonizadas. Para a recuperação de informação eficaz, faz-se necessário planejar uma estratégia de busca que se preocupe com as necessidades dos usuários. A estratégia de busca pode ser considerada “uma técnica ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada em uma base de dados” (LOPES, 2002a, p.61).

Em se tratando de estratégia de busca, as escolhas do banco/base de dados mais adequado, bem como do tipo de linguagem, se natural ou controlada, são exemplos de decisão que influenciam a eficácia do processo de recuperação da informação. Como linguagem natural (LN) entende-se a linguagem do discurso técnico-científico, denominado “texto livre”, ao passo que a linguagem controlada (LC) é o “conjunto limitado de termos autorizados para uso na indexação e busca de documentos”. Enquanto nas bases de dados verifica-se a LN nos campos de título e resumo, a LC encontra-se nos campos de descritores, termos de indexação ou identificadores (LOPES, 2002b, p.42).

PATENTES COMO FONTE DE INFORMAÇÕES TECNOLÓGICAS

As patentes constituem uma das formas mais antigas de proteção do capital intelectual. O século XIX caracterizou-se pela consolidação da proteção patentária em nível mundial. Segundo Barbosa (2003, p. 616), “o propósito da patente é incentivar a produção de novas tecnologias, através da garantia jurídica da exclusividade de seu uso”.

Os elementos constituintes do documento de patente são folha de rosto, relatório descritivo, desenhos (se houver), reivindicações e resumo. A folha de rosto apresenta os dados

formais da patente, tais como nome do(s) inventor(es), país de origem, nome do titular da patente, classificação internacional e o resumo. O relatório descritivo faz a descrição do objeto da invenção (produto e/ou processo) de modo a possibilitar a sua realização por um técnico no assunto. O teor das reivindicações, baseadas nas informações constantes do relatório descritivo, é o que determina a extensão da proteção conferida pela patente.

O sistema de patentes, segundo Macedo *et al.* (2001), fornece o alicerce para a reunião, a classificação e a disseminação de informação tecnológica, contribuindo para racionalização dos recursos empregados em pesquisa e desenvolvimento. Em um processo decisório, as informações constantes nas patentes auxiliam na antecipação dos atos e decisões dos concorrentes. A análise da documentação patentária, em determinado segmento tecnológico de interesse, permite que empresas e centros de P&D determinem o rumo dos investimentos e linhas de pesquisas que devem ser adotadas e evitam as “re-invenções”.

Parafraseando Rezende (2002), os agentes intérpretes, especialistas em análise e planejamento tecnológico, devem utilizar a informação como ferramenta de prospecção e identificação de novos negócios, mercados e tecnologias, a fim de identificar ameaças e oportunidades e antever mudanças de cenários. Segundo Castells e Bosch (2001), uma boa vigilância tecnológica deve permitir determinar:

- tecnologias pesquisadas publicadas em uma determinada área do conhecimento;
- soluções tecnológicas disponíveis;
- tecnologias emergentes;
- dinâmica das tecnologias (tecnologias em obsolescência e imponentes);
- linhas de pesquisa, equipes e líderes na geração de novas tecnologias.

Os documentos de patente possibilitam a concretização de todos os fatores supramencionados. A pesquisa em bancos de patentes nacionais é de interesse para os industriais brasileiros, tanto em áreas tecnológicas densamente protegidas, quanto naquelas em que há número reduzido de pedidos de patentes. Segundo Ahlert (1992), nas áreas com grande número de patentes, a preocupação principal seria a identificação de “espaços vazios”, ou seja, a procura do que ainda não foi pesquisado, ao passo que, nas áreas com baixa densidade de depósitos, pode-se procurar informação em bancos de patentes

estrangeiros, de modo a determinar a eventual disponibilidade de tecnologia não protegida no país.

Neste contexto, a recuperação da informação por meio dos bancos de patentes na *Web* torna-se fonte essencial de conhecimento. No entanto, as divergências observadas no processo de indexação de uma patente podem acarretar perda da informação e, conseqüentemente, decisões errôneas. Logo, um dos aspectos determinantes na recuperação da informação tecnológica perpassa pela necessidade de classificar o documento de maneira adequada.

CLASSIFICAÇÃO E CATEGORIZAÇÃO DOS DOCUMENTOS

A Classificação Internacional de Patentes (Int. Cl.) foi implementada pelo Acordo de Estrasburgo (1971) e entrou em vigor no Brasil em 1975, por meio do Decreto n° 76.472. A classificação foi criada com o objetivo de uniformizar a sistematização dos documentos de patente de invenção e servir como ferramenta de busca eficaz para a recuperação destes documentos por usuários do sistema de proteção patentária. O texto da Classificação Internacional é revisado periodicamente com o intuito de contemplar o desenvolvimento técnico-científico.

A classificação representa todo o conhecimento que possa ser considerado apropriado ao campo das invenções e está dividida em seções, classes, subclasses, grupos e subgrupos. Assim, um símbolo completo da classificação compreende os símbolos combinados que representam a classe, subclasse e o grupo principal ou subgrupo. Os subgrupos são constituídos pelo símbolo da subclasse, seguido de um número com um a três dígitos do seu grupo principal, da barra oblíqua e de um número com pelo menos dois dígitos que não seja 00, como, por exemplo, A61K 9/02 (supositórios ou similares). A estrutura hierárquica da classificação é dada por uma distinção da tecnologia em diversos níveis, ou seja, de sua divisão (seção, classe, subclasse, grupo e subgrupo) em ordem decrescente de hierarquia. A figura 1 apresenta o documento classificado como A61K 9/00.

A classificação internacional é um *tesauro* (termo autorizado: descritor e não descritor, de um domínio particular do conhecimento tendo em vista relações semânticas e lógicas, sendo sua construção apoiada no *corpus* discursivo da área

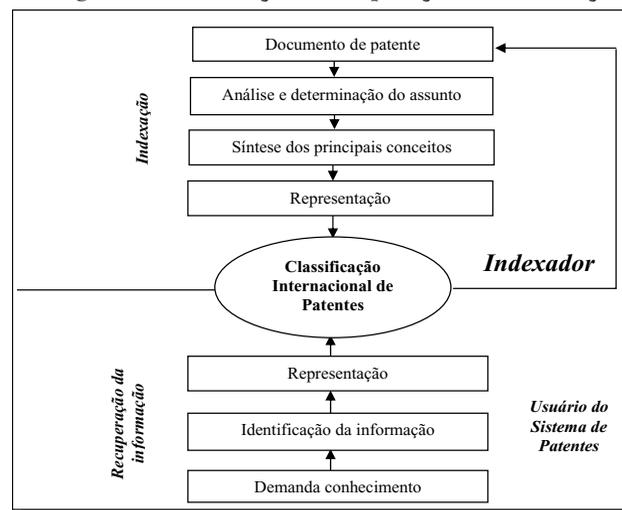
FIGURA 1
Exemplo de classificação

A	- Seção	→	Necessidades Humanas
61	- Classe	→	Preparações para finalidades médicas, odontológicas ou de toalete
K	- Subclasse	→	Preparações farmacêuticas
9/00	- Grupo	→	Caracterizadas pela forma

que abrange) que serve como guia para que os indexadores e usuários do sistema de patentes representem em uma linguagem comum os diversos campos tecnológicos. O ato de classificar uma matéria técnica de invenções perpassa pelas etapas de análise do documento, identificação dos seus principais conceitos e representação destes conceitos em uma linguagem de indexação.

A classificação foi projetada de modo que uma matéria técnica seja classificada e possa ser recuperada de um mesmo local dentro da Int. Cl. que provavelmente será examinado na busca dessa matéria. No campo das invenções tecnológicas, é mister garantir que a matéria seja categorizada adequadamente a fim de subsidiar a recuperação da informação. O processo de indexação e recuperação da informação de uma patente é mostrado na figura 2.

FIGURA 2
Fluxograma da indexação e recuperação da informação



O indexador (documentalista) deve verificar a informação contida no documento e sintetizá-la para que seja feita a representação baseada na Classificação Internacional de Patentes. Segundo Galvão (1998), a representação de um documento é temática ou descritiva. A representação temática é realizada por uma linguagem documentária que permite maior ou menor grau de abstração e generalizações. A representação descritiva contempla dados editoriais do documento que, no caso das patentes, referem-se ao titular da patente, inventor(es), país de origem, título da patente, número do documento etc.

Em contrapartida, os usuários do sistema patentário – pesquisadores, empresas e inventores isolados – buscam informações estratégicas no conhecimento revelado pelos documentos de patentes. Assim, é

necessário identificar com precisão o que o usuário deseja saber, a fim de que a representação da demanda seja condizente com suas expectativas. A interface entre indexador e usuário do sistema de patentes é dada pela Classificação Internacional de Patentes, que representa uma linguagem comum para o acesso aos documentos. Todavia, a indexação dos documentos faz uso de critérios subjetivos, uma vez que os indivíduos determinam, de forma independente, os objetos selecionados para a construção de suas representações.

Nota-se que a individualização perpassa pelo processo de categorização, dado que as descrições definidas distinguem uma classe única de conhecimento mediante a análise lógica das classes. Se quisermos classificar, por exemplo, o “uso do estrogênio como medicamento para o tratamento da menopausa”, tem-se que a classe dos medicamentos se intercepta com a classe dos tratamentos para menopausa. Neste cruzamento, identifica-se um só descritor, o estrogênio. Os nomes próprios permitem designar a mesma coisa no tempo e no espaço, singularizando um descritor e o diferenciando dos demais. Dessa forma, ao nomear *shampoo* para um determinado cosmético, excluem-se os demais produtos de mesma classe (creme rinse, batom, pó compacto, sabonete etc.). Este processo de construção da identidade do documento em análise é determinativo na fixação da Classificação Internacional.

A dificuldade de aplicação da Classificação Internacional de Patentes fez com que alguns escritórios de patentes recorressem a sistemas de classificação paralelos, dado que a Int. Cl. tornou-se inadequada às necessidades de alguns países. Este é o caso, por exemplo, do Escritório de Patentes Americano (USPTO), que, apesar de publicar seus documentos de patentes com a respectiva Classificação Internacional, também exhibe, na folha de rosto do documento de patente, a classificação estadunidense.

O advento da tecnologia da informação fez com que os métodos de revisão e emprego da classificação fossem reanalisados de modo a garantir sua eficiência e eficácia. O plano estratégico da reforma da classificação aprovado pelo Comitê de Especialista da Int. Cl. da Organização Mundial da Propriedade Intelectual em março de 2000, de acordo com Makarov (2004), teve como principais considerações:

- a classificação internacional deve ser aplicada mundialmente para qualquer tipo de invenção tecnológica;
- uma única classificação deve ser usada como base para indexação dos documentos de patentes;

- a edição da Int. Cl. corrente deve ser capaz de proporcionar uma representação adequada para a busca de documentos de patentes;
- a Int. Cl. deve se tornar uma ferramenta prontamente disponível e de fácil entendimento e uso;
- a classificação deve satisfazer as necessidades dos usuários do sistema de patentes sem distinção;
- A Int. Cl. deve ser compatível com meios eletrônicos.

DIVERGÊNCIAS NA INDEXAÇÃO DOS DOCUMENTOS DE PATENTES

A Classificação Internacional de Patentes enquadra-se na tipologia de classificação baseada na propriedade qualificada como diferença específica, apresentando-se como uma classificação mais complexa do que as chamadas dicotômicas. Como classificação documental, tem caráter pragmático, ou seja, foi concebida privilegiando mais o seu uso do que o domínio que vem a representar, conforme posto por Apostel (1963, *apud* POMBO, 2002).

Apesar dos esforços de se tentar estabelecer uma classificação internacional de patentes capaz de uniformizar a sistematização dos documentos de patente de invenção e servir como ferramenta de busca eficaz para a recuperação destes documentos por usuários do sistema de proteção patentária, verificam-se ainda discrepâncias no que se refere aos processos de indexação e recuperação da informação. Essas discrepâncias são atribuídas à inexistência de uma base cognitiva comum entre os profissionais responsáveis pelo tratamento da informação que farão a classificação e indexação do documento nos sistemas de informação e os usuários desses sistemas, correspondendo respectivamente aos segundo e terceiro estágios cognitivos definidos por Alvarenga (2003).

Tais discrepâncias fazem com que uma mesma matéria técnica seja classificada em locais distintos, o que caracteriza um comportamento oposicionista às premissas da criação da Classificação Internacional. Divergências na indexação dos documentos também estão relacionadas ao fato de que o conhecimento tácito é de difícil transformação naquele que pode ser representado pela linguagem (conhecimento explícito).

No processo de construção da identidade dos documentos de patentes, é evidente a manifestação das idéias ou prioridades do indexador (subjetividade) que representa o documento por meio de um modelo de ordenação do conhecimento de acordo com as experiências. Devido às especializações profissionais dos indivíduos, cada um apresenta estilo cognitivo próprio e

preferência por metodologias e ferramentas com as quais esteja familiarizado (SILVA, 2002). Ainda que se busquem padronizações no modo de pensar e agir humano, o estoque de conhecimento das pessoas é único, e somente o indivíduo pode determinar que fatos e circunstâncias devem ser levados em consideração no ato de indexação de um documento de patente.

A coerência da indexação pode ser facilitada pelo uso de vocabulários controlados, como é o caso da Int. Cl., mas o uso de termos livres, na opinião de Araújo Junior (2007), deve estar associado no processo de indexação, a fim de buscar alternativas para a representação temática dos conteúdos dos documentos para possibilitar a adaptação de termos que não estejam previstos dentre os instrumentos da linguagem controlada. Lopes (2002b) reforça essa opinião, dizendo que, ao analisar os resultados de diversos estudos sobre o uso de LN e LC na recuperação da informação, verificou que esses têm se concentrado na utilização conjunta dos dois tipos de linguagem na estratégia de busca, verificando-se que o uso simultâneo dessas linguagens proporciona melhor desempenho nos resultados.

Outro ponto a ser ressaltado reside no fato de que, segundo Schutz (1979), o estoque de conhecimento existe em um fluxo contínuo e está sujeito a constantes mudanças. Desta maneira, o que um indivíduo conhece hoje é diferente do conhecimento de amanhã, dado que o ser humano é capaz de aprender a aprender com suas experiências por meio da percepção do mundo exterior. Esse fato influencia o processo de categorização dos documentos. Como os indivíduos são capazes de ampliar e reestruturar seus conhecimentos continuamente, a interpretação dos conceitos destes documentos sofre mutações e, por conseguinte, há uma alteração nos mecanismos de representação da informação.

Os processos de percepção, aprendizagem, memória e pensamento produtivo que auxiliam na elaboração dos conceitos de um indivíduo são transformados a cada instante. Assim, a construção do conhecimento é ininterrupta, o que torna impossível padronizar o ato humano de indexação dos documentos. Ilustrativamente, a tabela 1 mostra um exemplo no qual o mesmo documento de patente foi indexado no Brasil e nos Estados Unidos em locais distintos da Classificação Internacional de Patentes.

Ainda que a Classificação Internacional de Patentes seja uma tentativa de mapeamento do

conhecimento e do estabelecimento de um vocabulário controlado no campo das invenções, os usuários do sistema não compartilham de uma linguagem comum com os inventores e indexadores. Os autores de patentes contribuem para a ampliação desta ausência de termos mínimos necessários à codificação do conhecimento, uma vez que os inventores são livres para descrever, da maneira que lhes é conveniente, suas invenções, inclusive criando neologismos. Desta forma, há uma multiplicidade de sinônimos que dificulta a busca de informações por palavras-chave, bem como a indexação dentro da classificação.

A representação do conhecimento, descrita pelo conjunto de convenções sintáticas e semânticas, sofre influência do contexto integrado por cada indivíduo (eu) que deve entrar em consonância para criar uma consciência coletiva unificada (nós), de modo que se tenha uma compatibilidade da linguagem entre a terminologia do indexador, a terminologia do usuário do sistema de proteção patentária e a terminologia da Classificação Internacional de Patentes.

As divergências na classificação dos documentos advindas de estratégias cognitivas distintas, além de tornar a tarefa de recuperação da informação laboriosa, impedem a identificação de ameaças concorrenciais e oportunidades de investimento em pesquisa e desenvolvimento pelos usuários do sistema de proteção patentária.

CONCLUSÃO

A subjetividade na indexação dos documentos de patentes gera divergências na Classificação Internacional atribuída às patentes de invenção e dificulta o acesso à informação tecnológica relevante para os gestores de inovações. Fatores determinantes na divergência de categorização dos documentos de patentes incluem a ausência de uma base

TABELA 1
Divergências na indexação de documentos de patentes

Nº do documento	BR 9508124-0	US 6,126,956
Int. Cl.	A61F 2/00	A61K 47/32
Descrição	Filtros implantáveis nos vasos sanguíneos.	Preparação medicinal caracterizada pelos ingredientes não ativos.
Nº do documento	BR 9500206-5	JP 7206680
Int. Cl.	A61K 31/5415	C07D 417/12
Descrição	Preparações medicinais contendo ingredientes ativos orgânicos.	Compostos heterocíclicos contendo dois ou mais heteroanéis (classificado pelo princípio ativo – piroxicam).
Nº do documento	BR 9501827-1	EP 679587
Int. Cl.	A61J 1/03	B65D 75/36
Descrição	Recipiente adaptado para finalidades farmacêuticas – pílulas ou tabletes.	Embalagem de bolhas (“blister”).
Nº do documento	BR 9502047-0	US 2001/0021728
Int. Cl.	A61C 13/00	A61F 2/00
Descrição	Próteses dentárias e métodos de fabricação.	Próteses, isto é, substitutos artificiais ou substituições de parte do corpo.

cognitiva comum entre os indexadores e usuários do sistema de proteção patentária, a dificuldade de transformação do conhecimento tácito em conhecimento explícito e a capacidade de os indivíduos renovarem e reestruturarem, constantemente, seus conhecimentos por meio da interação com o mundo exterior.

A linguagem utilizada como pilar da recuperação da informação apresenta uma dicotomia em relação à categorização dos documentos de patentes. Se, por um lado, é útil para a construção da identidade, por outro, dá lugar à imprevisibilidade na interpretação desses documentos, o que leva a representações díspares. Destarte, deve-se considerar a questão dos indexadores como ponto crucial da subjetividade existente na indexação. Destaca-se que a abordagem de que a Classificação Internacional deve satisfazer às necessidades dos usuários sem distinção reflete a indispensabilidade do estabelecimento de um padrão representativo mínimo do modelo mental do usuário para aproximá-lo do sistema de proteção patentária que seria, então, utilizado em toda a sua plenitude.

Artigo submetido em 15/11/2007 e aceito em 03/04/2008.

REFERÊNCIAS

- AHLERT, I. B. Agente x desenvolvimento. *Revista da ABPI*, Rio de Janeiro, n. 4, p. 27, 1992.
- ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. *Encontros Bibli*, Florianópolis, n. 15, 2003.
- ARAÚJO, C. A. A. A ciência da informação como ciência social. *Ciência da Informação*, Brasília, v. 32, n. 3, p. 21-27, 2003.
- ARAÚJO JUNIOR, R. H. *Precisão no processo de busca e recuperação da informação*. Brasília: Thesaurus, 2007.
- BALLONE, G. J. Conceitos e realidade. *PsiquWeb*. Disponível em: <<http://virtualpsy.locaweb.com.br/index.php?sec=47&art=264>>. Acesso em: maio 2006.
- BARBOSA, D. B. *Uma introdução à propriedade intelectual*. 2. ed. Rio de Janeiro: Lumen Juris, 2003.
- BORGES, M. E. N. et al. Estudos cognitivos em ciência da informação. *Encontros Bibli*, Florianópolis, n.15, p. 1-17, 2003.
- BRAGA, G. M. Informação, ciência da informação: breves reflexões em três tempos. *Ciência da Informação*, Brasília, v. 24, n. 1, 1995.
- CASTELLS, P. E.; BOSCH, R. M. La vigilancia tecnológica, requisito imprescindible para la innovación. In: _____. *El arte de innovar en la empresa*. 1. ed. Barcelona: Ediciones Del Bronce, 2001. p. 97-132.
- GALVÃO, M. C. B. Construção de conceitos em ciência da informação. *Ciência da Informação*, v. 27, n. 1, p. 46-52, 1998.
- GARDNER, H. *A nova ciência da mente*. São Paulo: Editora da Universidade de São Paulo, 1996.
- LIMA, G. A. B. Interfaces entre ciência da informação e ciência cognitiva. *Ciência da Informação*, Brasília, v. 32, n. 1, p. 77-87, 2003.
- LOPES, I. L. Estratégia de busca na recuperação da informação: revisão da literatura. *Ciência da Informação*. Brasília, v. 31, n. 2, p. 60-71, 2002a.
- _____. Uso das linguagens controlada e natural em base de dados: revisão da literatura. *Ciência da Informação*, Brasília, v. 31, n. 1, p. 41-52, 2002b.
- LUCAS, C. R. A metalinguagem como lugar da interpretação: terminologia e bases de dados informatizadas. *DELTA*, v. 15, n. 1, p.00-00, 1999.
- MACEDO, M. F. G.; MÜLLER, A. C. A.; MOREIRA, A. C. *Patenteamento em biotecnologia*. Brasília: Embrapa Comunicações para Transferência de Tecnologia, 2001.
- MAKAROV, M. The process of reforming the International Patent Classification. *World Patent Information*, v. 26, p. 137-141, 2004.
- POMBO, O. Da classificação dos seres à classificação dos saberes. *Leituras. Revista da Biblioteca Nacional de Lisboa*, n. 2, p. 19-33, 1998.
- REZENDE, Y. Informação para negócios: os novos agentes do conhecimento e a gestão do capital intelectual. *Ciência da Informação*, Brasília, v. 31, n. 2, p. 120-128, 2002.
- SILVA, S.L. Informação e competitividade. *Ciência da Informação*. Brasília, v. 31, n. 2, p. 142-151, 2002.
- SHCUTZ, A. O mundo da vida. In: HELMUT, R. W. (Org.). *Fenomenologia e relações sociais*. Rio de Janeiro: Zahar, 1979.
- TRISTÃO, A. M. D.; FACHIN, G. R. B.; ALARCON, O. E. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. *Ciência da Informação*, Brasília, v. 33, n. 2, p. 161-171, 2004.