

NORMA PORTUGUESA

NP 3715

1989

Documentação

Método para a análise de documentos, determinação do seu conteúdo e selecção de termos de indexação

Documentation

Méthodes pour l'analyse des documents, la détermination de leur contenu et la sélection des termes d'indexation

Documentation

Methods for examining documents, determining their subjects and selecting indexing terms

CDU

001.815

Descritores

Ciência da documentação e informação; descrição do documento; indexação; linguagens de indexação; thesauri

Correspondência

ISO 5963, equivalente

Homologação

Diário da República, III Série, N.º 186, de 1989-08-14

Elaborado por

CT7 (IPQ)

Edição

Outubro de 1989

© IPQ reprodução proibida

ERRATA
(Fevereiro de 1992)

NP 3715(1989)

Documentação
Método para análise de documentos, determinação do
seu conteúdo e selecção de termos de indexação

Onde se lê:	Leia-se:
P. 1, título em francês déterminaion	détermination
P. 3, secção 1.3, 6a linha serem idênticas	não serem idênticas
P. 8, secção 7.4, 1a linha indicaponderação	indicadores de função, de liga- ções, de ponderação, etc.
P. 8, secção 8.3, 2a linha enteder	entender
P. 10, Anexo amguidade	ambiguidade

1. Objectivo e campo de aplicação

1.1. A presente Norma recomenda procedimentos para a análise de documentos, determinação do seu conteúdo e selecção de termos de indexação. O seu objectivo limita-se a estas primeiras fases de indexação e pretende ser independente das práticas de qualquer sistema de indexação, quer seja pré ou pós-coordenado. Descreve os métodos gerais para análise de documentos que devem aplicar-se a qualquer situação.

Estes métodos, contudo, destinam-se especialmente aos sistemas de indexação em que se exprime o conteúdo dos documentos de forma resumida, com a ajuda dos termos de uma linguagem de indexação controlada que representa as noções pretendidas. Neste contexto, a linguagem controlada é geralmente constituída por um subconjunto de termos extraídos da linguagem natural, organizados, por exemplo, num thesaurus. No entanto, podem aplicar-se estes métodos a sistemas em que os conceitos são representados, para efeito de pesquisa documental, por símbolos escolhidos num plano de classificação.

1.2. As técnicas descritas na presente Norma podem ser empregues em qualquer organismo onde a análise dos documentos e a expressão do seu conteúdo em termos de indexação sejam feitas por indexadores. Não se aplicam aos organismos que utilizam técnicas de indexação automática nas quais os termos de um texto são organizados em conjuntos ou classes, segundo critérios que podem ser aplicados por computador, como, por exemplo, a frequência da sua utilização e/ou a adjacência no texto, ainda que a finalidade deste sistema seja a mesma.

1.3. A presente Norma deve, em primeiro lugar, servir de guia aos indexadores para as fases da análise dos documentos e identificação dos conceitos. Pode também ser aplicada à pesquisa documental, transformando as perguntas dos utilizadores em termos de indexação controlados. Do mesmo modo, pode servir de guia para a elaboração de resumos analíticos. No entanto, deve chamar-se a atenção para o facto de estas tarefas, ainda que análogas, serem idênticas.

1.4. A presente Norma destina-se a promover a utilização de uma prática normalizada:

- a) num organismo ou numa rede de organismos;
- b) entre diferentes serviços de indexação, particularmente naqueles que permutam informações bibliográficas.

2. Referências

ISO 2788

Documentation. Principes directeurs pour l'établissement et le développement de thesaurus monolingues.

ISO 5964

Documentation. Principes directeurs pour l'établissement et le développement de thesaurus multilingues.

3. Definições

Para os fins da presente Norma, utilizam-se as seguintes definições:

3.1. Documento: qualquer unidade de informação impressa ou não, passível de catalogação ou indexação.

NOTA: Esta definição designa não só materiais escritos e impressos, em papel ou microforma (livros, revistas, mapas, diagramas, etc.), mas também formas de comunicação não impressas (registos legíveis por máquina, filmes e registos sonoros) e objectos, por exemplo peças de colecção.

3.2. Conceito: unidade de pensamento. O conteúdo semântico de um conceito pode ser reexpresso combinando outros conceitos que podem ser diferentes de uma língua/cultura para outra.

3.3. Assunto: qualquer conceito ou combinação de conceitos que representem um tema num documento.

3.4. Termo de indexação: representação de um conceito sob a forma de um termo derivado da linguagem natural, de preferência um substantivo simples ou composto, ou de um símbolo de notação de uma classificação.

NOTA: Um termo de indexação pode ser constituído por mais de uma palavra. Numa linguagem de indexação controlada, um termo pode ser designado por descritor ou não descritor.

3.5. Descritor: termo que se utiliza na indexação para representar um determinado conceito, por vezes chamado termo preferencial.

3.6. Não descritor: sinónimo ou quase sinónimo de um descritor. Não pode ser atribuído a documentos mas serve de entrada num thesaurus, sendo, neste caso, o utilizador remetido, através de uma nota (por exemplo «ver» ou «usar»), para o descritor apropriado; por vezes chamado termo não preferencial.

3.7. Índice: lista alfabética e sistemática de assuntos que remetem, cada um deles, para a respectiva localização nos documentos.

3.8. Indexação: acção que consiste em descrever ou caracterizar um documento relativamente ao seu conteúdo, representando esse conteúdo numa linguagem documental.

4. Procedimentos de indexação

4.1. A indexação não consiste na descrição de um documento enquanto entidade física (por exemplo, não indica a forma, o editor, a data, etc.). De qualquer modo, estes elementos podem ser incluídos num índice de matérias se permitirem a um utilizador determinar com mais precisão se um dado documento constitui ou não resposta à sua questão.

4.2. Na indexação, extraem-se os conceitos dos documentos, por um processo de análise intelectual, sendo, seguidamente, convertidos em termos de indexação. A análise e a conversão devem ser feitas recorrendo a instrumentos de indexação tais como thesauri e planos de classificação.

4.3. A indexação divide-se essencialmente em três fases que, na prática, tendem a sobrepor-se:

- a) análise do documento e definição do seu conteúdo;
- b) identificação e selecção dos conceitos representativos do conteúdo;
- c) representação desses conceitos por termos de indexação.

Cada uma destas fases assim como o controlo da qualidade são tratados mais adiante nos números 5 a 8.

5. Análise do documento

5.1. A precisão com que um documento pode ser analisado depende em grande medida da sua forma física. Podem distinguir-se dois casos: os documentos escritos e os documentos não escritos.

5.2. Os documentos escritos (geralmente impressos) constituem o fundo habitual das bibliotecas e dos centros de documentação, integrando monografias, periódicos, relatórios, actas de congressos, etc. A apreensão total destes documentos implica, em princípio, uma leitura minuciosa dos textos. Nem sempre é necessária uma leitura completa, contudo, o indexador deve certificar-se de que nenhuma informação útil foi descurada. As partes importantes do texto devem ser analisadas com cuidado, devendo ser dada uma especial atenção aos seguintes elementos:

- a) título;
- b) resumo, se houver;
- c) sumário;
- d) introdução, início dos capítulos e parágrafos e conclusão;
- e) ilustrações, diagramas, quadros e respectivas legendas;
- f) palavras ou grupos de palavras sublinhadas ou realçadas por um tipo de letra diferente.

Todos estes elementos devem ser cuidadosamente examinados e registados pelo indexador, quando analisa o documento. Não é recomendada uma indexação feita com base unicamente no título e, se houver um resumo, este não deverá substituir a análise do texto. Com efeito, alguns títulos podem induzir em erro; certos resumos são insuficientes e nem uns nem outros constituem uma fonte segura para o tipo de informação de que o indexador necessita.

5.3. Os documentos não escritos, tais como os documentos audiovisuais, visuais ou sonoros, incluindo objectos, requerem procedimentos diferentes. Na prática, nem sempre é possível analisar um registo na sua totalidade (por exemplo, a projecção de um filme). A indexação é habitualmente feita com base no título ou no resumo, tendo o indexador sempre a possibilidade de ver ou ouvir o documento se a descrição for insuficiente ou parecer inexacta.

6. Identificação e selecção dos conceitos

6.1. Após análise do documento, o indexador deve identificar os conceitos representativos do seu conteúdo. Os organismos devem construir grelhas de identificação que contenham os critérios considerados importantes na área abrangida pela indexação.

As questões seguintes constituem exemplos de critérios que essas grelhas devem fixar:

- a) o documento ocupa-se de um objecto afectado por uma actividade?
- b) O assunto contém um conceito (por exemplo uma acção, uma operação, um processo, etc.)?
- c) O objecto é afectado pela actividade identificada?
- d) O documento ocupa-se do agente desta acção?
- e) Tem necessidade de recorrer a meios especiais para realizar a acção (por exemplo instrumentos, técnicas ou métodos especiais)?
- f) Estes elementos estão inseridos no contexto de um lugar ou meio ambiente particulares?

g) As variáveis dependentes ou independentes estão identificadas?

h) O assunto foi tratado de um ponto de vista particular que normalmente não se associa a esta área de estudo (por exemplo um estudo sociológico da religião)?

Estes são exemplos de critérios susceptíveis de serem aplicados a muitas áreas.

Pode ser necessário formular outras perguntas para construir uma grelha de indexação em determinadas disciplinas.

6.2. O indexador não tem necessariamente de considerar como termos de indexação todos os conceitos identificados ao analisar o documento. A escolha destes conceitos depende do fim para o qual os termos de indexação vão ser utilizados.

Esse fim pode variar entre a produção de índices alfabéticos impressos e a constituição de um banco de dados bibliográficos informatizado. A identificação dos conceitos pode também ser afectada (como foi referido anteriormente) pelo documento indexado. Por exemplo, a indexação de livros pode ser diferente da indexação de artigos de periódicos. Assim, as duas características de uma indexação mais susceptíveis de serem modificadas são a exaustividade e a especificidade.

6.3. A exaustividade está ligada ao número de noções que foram consideradas e que caracterizam o conteúdo integral do documento.

6.3.1. Um indexador que segue os procedimentos anteriormente indicados deve poder identificar, num documento, todos os conceitos que têm um valor potencial para os utilizadores de um sistema de informação. Em certos casos, aparecem no mesmo documento dois ou mais temas independentes um do outro, na área abrangida pela indexação. Devem ser tratados separadamente e, se necessário, por diferentes especialistas do assunto em questão.

6.3.2. O âmbito da indexação não deve ser limitado de forma demasiado restrita. Com o desenvolvimento das redes de informação, torna-se necessário chamar a atenção para o facto de os termos de indexação, inicialmente criados por um grupo de utilizadores tais como cientistas ou técnicos, poderem, com utilidade, ser estudados por outros grupos de utilizadores (por exemplo economistas). Também se recomenda aos indexadores de bibliografia científica e técnica que não descurem as outras facetas de um assunto, por exemplo o seu aspecto social ou económico.

6.3.3. O critério principal na selecção dos conceitos deve ser sempre o seu valor potencial, como um elemento na expressão do conteúdo de um documento e na sua recuperação. Ao escolher os conceitos, o indexador deve ter presentes as perguntas que podem ser feitas ao sistema de informação, tanto quanto possam ser conhecidas. Com efeito, este critério constitui a principal função da indexação. Neste contexto, o indexador deve:

a) escolher os conceitos mais apropriados a um determinado grupo de utilizadores, não perdendo de vista o objectivo da indexação;

b) modificar, se necessário, quer os instrumentos de indexação quer os procedimentos, a partir do resultado das perguntas colocadas ao sistema. Estas modificações não devem, porém, provocar uma distorção da estrutura ou lógica da linguagem de indexação.

6.3.4. O número de termos ou de descritores que pode ser atribuído a um documento não deve ser limitado de modo arbitrário. Deve ser inteiramente determinado pela quantidade da informação contida no documento em ligação com as supostas

necessidades dos utilizadores. Impor um limite arbitrário pode conduzir a uma perda de objectividade na indexação e a uma deformação da informação que pode ser útil para a recuperação. Se for necessário limitar o número de termos num determinado serviço, a selecção dos conceitos deve ser feita somente por decisão do indexador relativamente ao papel de cada conceito na expressão da totalidade do assunto do documento.

6.4. A especificidade está ligada à exactidão com que um determinado documento é representado por um termo de indexação. Pode ocorrer uma perda de especificidade quando um determinado conceito é representado por um termo que tem um significado mais geral.

Os conceitos devem ser identificados do modo mais específico possível. Podem preferir-se conceitos mais gerais nas seguintes circunstâncias:

- a) quando o indexador considere que um excesso de especificidade pode actuar de modo negativo no funcionamento do sistema de indexação. Um indexador pode decidir, por exemplo, que modelos muito específicos de equipamento podem ser representados por termos mais gerais, tais como o nome do fabricante, e talvez pela família dos modelos, em especial quando estas noções aparecem somente nas zonas limites da área abrangida pela indexação;
- b) segundo a importância atribuída a um conceito pelo autor, se o indexador considerar que uma ideia não está completamente desenvolvida, ou que só ocasionalmente lhe é feita uma alusão pelo autor, pode justificar-se uma indexação a nível mais geral.

7. Escolha dos termos de indexação

7.1. Ao traduzir os conceitos em termos de indexação, o indexador deve observar as seguintes regras (veja também o anexo):

- a) os conceitos já contidos na linguagem de indexação devem permanecer sob a sua forma preferencial;
- b) os termos que representam conceitos novos devem ser verificados quanto ao conteúdo e à forma, com a ajuda de instrumentos de referência tais como: dicionários e enciclopédias de reconhecida autoridade nas respectivas áreas; thesauri, nomeadamente os construídos de acordo com a ISO 2788 ou ISO 5964; planos de classificação.

Podem também ser consultados especialistas do assunto, particularmente aqueles que têm alguns conhecimentos nas áreas de indexação e documentação.

7.2. O indexador deve conhecer os instrumentos de indexação e ter presente que estes instrumentos podem impor determinadas limitações. Por exemplo uma lista de encabeçamentos de assuntos ou as rubricas de um plano de classificação podem não permitir uma representação exacta do conceito encontrado no documento. Se os conceitos são representados por códigos de classificação, é necessário saber que estes códigos designam, regra geral, uma classe mais genérica ou mais específica, que pode não ser totalmente aplicável ao documento em questão.

7.3. Se a linguagem de indexação está contida num thesaurus, a multiplicidade dos termos atribuídos a um documento e o número de entradas podem ser reduzidos sem prejuízo, visto que as relações genéricas e outras relações a priori podem ser estabelecidas directamente a partir do próprio thesaurus. Quando se utiliza um

thesaurus é necessário seleccionar o termo mais específico para representar um dado conceito.

7.4. Alguns sistemas de indexação utilizam indicaponderação, etc. O indexador deve estar familiarizado com todas estas regras particulares de sintaxe.

7.5. Na prática, o indexador encontrará frequentemente conceitos que não existem em nenhum thesaurus ou plano de classificação. De acordo com o sistema utilizado, estes conceitos podem ser tratados de várias maneiras, como por exemplo:

- a) através de termos ou descritores imediatamente aceites pela linguagem de indexação;
- b) representados temporariamente por termos mais genéricos, os novos conceitos ficam propostos como candidatos, para integração posterior.

8. Controlo da qualidade

8.1. A qualidade e a coerência da indexação dependem de determinados factores, tais como:

- a) a competência do indexador;
- b) a qualidade dos instrumentos de indexação.

Numa situação ideal, os termos de indexação atribuídos a um documento e o grau de exaustividade atingido na altura da indexação devem ser idênticos, independentemente do indexador. Estes factores devem, além disso, manter-se relativamente estáveis no tempo para um mesmo sistema de indexação. A coerência é um factor importante no funcionamento de um sistema eficaz de indexação, sobretudo quando a permuta da informação se realiza entre os diferentes centros de uma rede documental.

8.2. A total imparcialidade do indexador é um factor necessário para se obter a coerência da indexação. A subjectividade na identificação dos conceitos e a escolha dos termos de indexação afectarão inevitavelmente o funcionamento do sistema de indexação. A coerência é mais difícil de obter dentro de uma grande equipa de indexação ou quando esta é feita por equipas de indexadores que trabalham em lugares diferentes, tal como num sistema descentralizado. Em tais situações, recomenda-se uma etapa de verificação centralizada, com retroacção para os indexadores.

8.3. O indexador deve possuir um adequado conhecimento da área abrangida pelos documentos que indexa. Deve entender os termos encontrados nos documentos assim como as regras e procedimentos da linguagem de indexação específica. Os organismos que tratam documentos em línguas estrangeiras devem recorrer a especialistas das línguas em questão.

8.4. A qualidade da indexação será tanto melhor quanto mais directo for o contacto dos indexadores com os utilizadores. Poderão então determinar se certos termos ou descritores são susceptíveis de produzir combinações falsas e de criar, assim, saídas não pertinentes.

8.5. A qualidade da indexação depende também da possibilidade de actualização da linguagem de indexação. É essencial que o sistema permita a introdução de novos termos na linguagem, ou modificações na terminologia, correspondendo assim às novas necessidades dos seus utilizadores.

8.6. A qualidade da indexação deve, se possível, ser avaliada através da análise dos resultados da recuperação, calculando, por exemplo, as taxas de exaustividade e de pertinência.

9. Referência à normalização internacional

Esta Norma baseia-se na ISO 5963 «Documentation. Méthodes pour l'analyse des documents, la détermination de leur contenu et la sélection des termes d'indexation», à qual é equivalente.

ANEXO Organograma do procedimento de indexação utilizando um thesaurus (este anexo não faz parte integrante da Norma.

