



XV Encontro Nacional de Tecnologia do Ambiente Construído

Avanços no desempenho das construções – pesquisa, inovação e capacitação profissional

12, 13 E 14 DE NOVEMBRO DE 2014 | MACEIÓ | AL

MINERAÇÃO DE TEXTO APLICADA A ARTIGOS DO ENCAC E ELACAC (1990-2013) RELACIONADOS A ACESSO SOLAR

TAMURA, Cíntia A. (1); KRÜGER, Eduardo L. (2); GUIMARÃES, André J. R. (3)

(1) Universidade Tecnológica Federal do Paraná (UTFPR), (41)8870-1002, e-mail: cintiatamura@gmail.com, (2) UTFPR, e-mail: ekruger@utfpr.edu.br, (3) Universidade Federal do Paraná, e-mail: andrejrg@gmail.com

RESUMO

Para que haja aquecimento e iluminação passivos de edificações é necessário o acesso solar. Para sua obtenção, latitude e orientação solar devem ser consideradas na concepção de planos urbanísticos. Caso contrário, podem-se ter ambientes úmidos, mal iluminados e frios, causando desconforto aos usuários. Neste sentido, estudos surgidos a partir de 1990 vêm ampliando a questão, com a inclusão dos efeitos da falta de acesso solar em processos biológicos não ligados à visão. Quando o acesso solar é escasso ou inexistente, indivíduos podem apresentar índices maiores de estresse, depressão e ansiedade. A falta de luz solar pode também influenciar a regulação de ritmos biológicos como o ciclo circadiano e a produção de hormônios. Além disso, as fontes luminosas artificiais predominantemente não possuem espectro similar ao da luz natural, necessário para a ocorrência de diversas funções orgânicas. Desta forma, a prática corrente de especificação de níveis de iluminação artificial baseados somente em conforto visual pode ser inadequada para atender os requisitos necessários para estimulação biológica, sugerindo a necessidade de que o tema seja mais explorado. Assim, o objetivo do trabalho foi realizar o levantamento dos termos mais empregados em publicações do ENCAC/ELACAC desde seu início (1990) até 2013, identificando os termos mais frequentes relacionados ao tema acesso solar, e verificando a posição do termo saúde em relação aos demais. Foi utilizada a técnica de mineração de texto aplicada a um banco de dados com todas as publicações do evento. Os resultados indicaram que a produção relacionada ao acesso solar é substancialmente menor quando comparada percentualmente com o total de publicações da população avaliada. Os termos mais empregados relacionados são conforto térmico, eficiência energética e iluminação. As publicações que relacionam acesso solar e saúde são incipientes, sugerindo que a relação acesso solar e saúde oferece vasto campo a ser explorado pela comunidade científica.

Palavras-chave: Mineração de texto; Acesso solar; Saúde.

ABSTRACT

The solar access is necessary for passive heating and illumination in buildings. Its obtainment requires considering the latitude and solar orientation in conceptions of urban plans. If it does not happen, it can cause humid and cold ambiences with poor illumination that leads discomfort to its users. Furthermore, since 1990, studies have been expanding the issue and including the effect of lack of solar access in biological processes not related to vision. When the solar access is scarce or nonexistent, individuals can present bigger indexes of stress, depression and anxiety. The lack of solar light can also affect the regulation of biological rhythms like circadian cycle and hormone production. Moreover, the artificial light sources predominantly do not have spectrum similar to natural light that is required to have many organic functions. Thereby the current practice in specification of artificial illumination levels based only on visual comfort can be inadequate to attend the requirements necessary for biological stimulation. This fact suggests the need for the subject to be further explored. This research aims to survey the most used terms in articles in ENCAC/ELACAC since its beginning (1990) until 2013, identifying the most frequent terms related to solar access, and verifying the position of term health in relation to the others. This work uses text mining techniques applied to a database containing all publications of event. The results indicate that the production related to solar access is substantially smaller when compared in percentage to total number evaluated of publications. The terms most used are related to thermal comfort, energy efficiency and

illumination. The publications related to solar access and health are inexpressive. This finding suggests that relationship between solar access and health offers a vast unexplored field to the scientific community.

Keywords: *Text mining; Solar Access; Health.*

1 INTRODUÇÃO

O irreversível processo mundial de verticalização e adensamento das cidades (PEDRAZA et al., 2013; RAMIRES, 2011) vem transformado a malha urbana, gerando aglomerados de assentamentos cada vez mais complexos (LOBACCARO; FRONTINI, 2014). Neste contexto, as obstruções e sombreamentos gerados pela interação entre prédios cada vez maiores e próximos entre si pode resultar na diminuição ou mesmo impossibilidade de obtenção de acesso solar no interior das edificações (LITTLEFAIR, 2001; ESCH, VAN; LOOMAN; BRUIN-HORDIJK, DE, 2012).

Desta forma, determinar os efeitos dinâmicos que a trajetória diária do Sol pode causar entre as edificações pode ser o meio para a obtenção de subsídios visando o estabelecimento de recomendações urbanas de ocupação, de forma que a possibilidade de obtenção de condicionamento passivo (aquecimento e iluminação), e do aproveitamento do potencial fotovoltaico em construções possam ser preservados (LOBACCARO; FRONTINI, 2014).

Entretanto, a importância da utilização de luz natural no interior de espaços arquitetônicos não reside somente na economia de energia, mas também no fato de que a maioria dos indivíduos demonstra preferência pela luz natural, tanto em ambientes domésticos quanto em edifícios comerciais (MESA; CORICA; PATTINI, 2011). A luz natural proporciona ambientes mais atrativos, e cria um ambiente de trabalho mais adequados, onde as atividades são realizadas de forma mais confortável durante o dia. (ibid., 2011). Porém, da mesma forma como níveis adequados de acesso solar podem proporcionar conforto e incremento à produtividade, sua presença insuficiente ou inexistente pode resultar em ambientes excessivamente úmidos, pouco iluminados e frios; este conjunto de fatores pode causar efeitos negativos na saúde humana, como desconforto térmico e visual (PRADO, 2006; SEONG et al., 2006). É importante destacar também que um crescente corpo de evidências sugerem a ligação da ausência de luz natural à doenças severas. A escassez de luz natural pode fazer com que indivíduos desenvolvam depressão, demência, distúrbios no ritmo circadiano, fratura de ossos, disfunções renais, enfraquecimento do sistema imunológico, entre outros (BOUBEKRI, 2008).

Além disso, as fontes luminosas artificiais predominantemente não possuem espectro magnético como o da luz natural, necessário para a ocorrência de diversas funções orgânicas (HECHT, 2012). Assim, a prática corrente de obtenção de níveis adequados de iluminação artificial baseados somente em conforto visual pode ser inadequada para atender aos requisitos necessários para estimulação biológica, sugerindo a necessidade de ampliação da abordagem ao tema.

Sob esta perspectiva, o objetivo do trabalho foi realizar o levantamento dos termos mais empregados em artigos de pesquisas nacionais relacionadas ao acesso solar, e verificar em que posição o termo saúde ocupa em relação aos demais. Desta forma, foi possível descobrir quais temas relacionados ao conceito de acesso solar têm sido mais explorados, e evidenciar quais são os que ainda demandam mais estudos pela comunidade científica.

Para isso, foi utilizada a técnica de mineração de texto. Adotou-se como população avaliada todos os trabalhos científicos publicados nos anais do ENCAC/ELACAC desde seu início (1990) até 2013, pela relevância do evento no cenário nacional na área de iluminação natural e artificial.

2 REVISÃO BIBLIOGRÁFICA

Nesta seção, são apresentados os conceitos essenciais relacionados conceito de mineração de texto, método utilizado para a quantificação da produção relacionada ao acesso solar na população selecionada (anais ENCAC e ELACAC de 1990 a 2013).

2.1 Mineração de texto

A mineração de texto é uma atividade multidisciplinar, cujo objetivo é obter conhecimento não explícito de dados textuais, apresentando-o de forma estruturada, coerente e concisa (FARO; GIORDANO; NIE; GAO; CAO, 2012). Segundo Aggarwal e Zhai (2012), o papel da mineração de texto vai além de simplesmente facilitar o acesso à informação, pois a identificação de padrões e de tendências que este possibilita pode auxiliar a análise e assimilação de informações, facilitando o processo de tomada de decisão, que pode assim se basear em resultados obtidos de experiências passadas. No caso de pesquisas científicas, os resultados obtidos com a mineração de textos podem auxiliar na identificação de áreas ainda não totalmente exploradas, permitindo o direcionamento de novas pesquisas a nichos do conhecimento ainda pouco explorados.

Para Gajzler (2010), o processo de mineração de textos é composto por vários estágios, que começam na definição dos objetivos e terminam com a análise do resultado obtido pela mineração. Para o autor, a conversão de documentos de texto para um formato que possibilite a aplicação de algoritmos de cálculo e análise é a parte central da mineração e é dividida em quatro estágios. (1) Transformação dos textos: quando se remove ou se exclui símbolos desnecessários dos documentos avaliados, convertendo-os a um texto sem formatação; (2) Identificação das palavras (tokenização): divide o documento do texto contínuo em palavras distintas. Além disso, remove caracteres indesejados como sinais de pontuação, separação silábica, marcações especiais e números, elementos que isolados agregam pouca informação (SOARES; PRATI; MONARD, 2008); (3) Redução ao núcleo (*stemming*): reduz as palavras à sua forma básica (radical). Segundo Soares, Prati e Monard (2008), o objetivo dos algoritmos de *stemming* é remover os prefixos e sufixos dos termos, reduzindo as formas variantes a uma forma comum chamada *stem*. Além disso, esta etapa elimina palavras sem significância para o assunto estudado; (4) Matriz de frequência: os *stems* são contabilizados tanto em todo o material analisado, como em cada documento individual. Para Fan, Wallace, Rich e Zhang (2006), esta etapa gera além da contabilização, uma categorização dos *stems* definidos no estágio anterior. Esta categorização geralmente é baseada em um dicionário de sinônimos, que identifica termos mais próximos e mais distantes, bem como os principais temas abordados no documento.

Soares, Prati e Monard (2008) também destacam que a ocorrência de palavras em sequência pode conter mais informações que as palavras isoladas. Por isso, o procedimento chamado n-grama, aquele que estabelece a frequência de palavras unidas com outras adjacentes, ou mesmo, mais distantes, exerce um importante papel na mineração de textos. Kumar e Bhatia (2013) afirmam que dados textuais, diferentemente de dados estruturados, são ambíguos e difíceis de serem processados. No entanto, a maioria dos dados gerados pela sociedade atual, resultados de meios de comunicação, transmissão de informações ou manifestação de opiniões, pertence ao formato textual. Por isso, a mineração de texto, como ferramenta que busca extrair significados, segundo um propósito específico, de um texto com linguagem natural, não só recebe um papel de destaque na obtenção de informações, como compõe um campo de pesquisa (ibid., 2013).

3 MÉTODO

Quanto à metodologia, a pesquisa é (1) básica, por sua natureza; (2) quantitativa, quanto à abordagem do problema; (3) exploratória, do ponto de vista dos objetivos; e (4) bibliográfica, quanto aos procedimentos técnicos. O processo de mineração de textos adotado envolveu a realização de quatro etapas: (1) Preparação do banco de artigos; (2) Processo de mineração de textos e; (3) Pré-análise dos termos minerados.

3.1 Preparação do banco de artigos

A população avaliada se constituiu em todos os artigos técnicos publicados no Encontro Nacional e Encontro Latino Americano de Conforto no Ambiente Construído (ENCAC / ELACAC). O encontro bienal, focado na área de Conforto Ambiental, envolve campos de estudo como Conforto Térmico, Iluminação Natural e Artificial, Acústica, Eficiência Energética, Ergonomia e Avaliação Pós-Ocupação. Ao longo de 22 anos, 12 encontros ocorreram, sendo o primeiro em 1990, e o último, em 2013.

Estas 12 edições resultaram em 2062 artigos (disponibilizado pela organização do evento aos participantes da última edição, 2013), que constituem o banco de dados, objeto de estudo da pesquisa. Originalmente disponibilizados em arquivos no formato PDF, os arquivos foram transformados para texto sem formatação por meio do *software* Simpo PDF to Text, versão 2.1.5 que converteu cada arquivo PDF para um arquivo com extensão TXT. Porém, durante este procedimento foram identificados alguns arquivos PDF ilegíveis que não puderam ser convertidos. Por este motivo, foram descartados da população 15 artigos, sendo nove da edição 2001, um de 2003, um de 2005 e quatro de 2013.

Ainda como preparação do banco de artigos, os 2047 arquivos convertidos para texto foram submetidos a um algoritmo PHP que retirou todos os caracteres especiais do seu conteúdo, como acentos, pontuação, tabulações e quebras de linhas.

Por fim, foi constatado que os arquivos da edição de 1990 do evento foram digitalizados, sendo que apenas o resumo pode ser extraído como texto. Assim, para estes casos, a mineração de textos só pode ser aplicada ao resumo. Para os arquivos das demais edições do evento, os artigos completos foram analisados.

3.2 Processo de mineração de textos

Para a mineração de textos, foi utilizado o *software* PreText 2 (LABORATORY OF COMPUTATIONAL INTELLIGENCE, 2013) para realizar os procedimentos de (1) tokenização: divisão de texto contínuo em termos (*tokens*), removendo caracteres indesejados como sinais de pontuação, separação silábica, marcações especiais e números; (2) *stemming*: redução de cada *token* para o radical que o originou, utilizando algoritmos de *stemming*; (3) remoção de *stopwords*: descarte de palavras que são muito comuns no idioma utilizado, como pronomes, artigos, preposições, advérbios, conjunções, e que, em princípio, não são relevantes para a caracterização do texto analisado; e (4) análises n-grama: análise dos termos em sequência, por meio da união de duas ou mais palavras consecutivas.

Todo o procedimento realizado com o PreText 2 foi parametrizado de forma a retornar *stems* que estivessem presentes em, pelo menos, cinco artigos, com o intuito de reduzir a diversidade de *stems* encontrados. Além disto, as análises foram limitadas aos 20 primeiros *stems* que mais ocorreram.

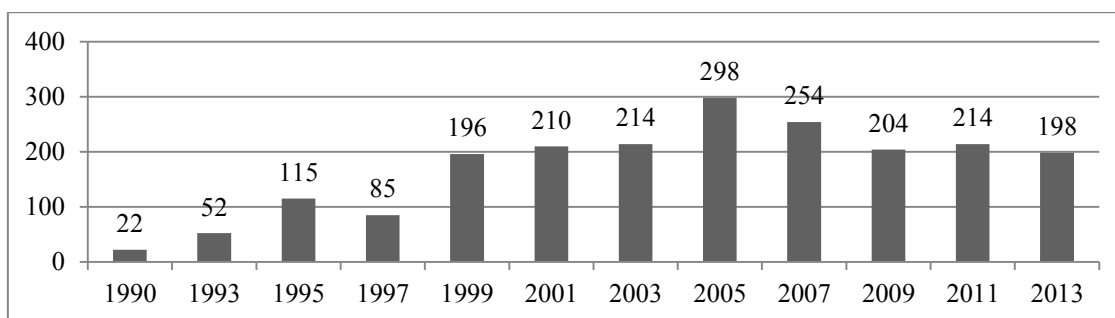
3.3 Pré-análise dos termos minerados

O resultado obtido pelo PreText 2 gerou quatro grandes listas de termos, sendo a primeira com os *stems* mais recorrentes em todos os artigos analisados, e as demais referentes às análises 2-grama, 3-grama e 4-grama. Para identificação dos 20 termos mais relevantes para a pesquisa, bem como a posição do primeiro termo relacionado à saúde, foram identificados e retirados os *stems* referentes a expressões comuns em produções científicas, mas que pouco acrescentam ao tema estudado (exemplo: ‘objetivo’, ‘pesquisa’, ‘figura’, ‘tabela’, ‘resultado’, ...). Nas análises referentes à junção de mais de *stem*, foi alta a frequência de expressões referentes a nomes de instituições acadêmicas e do próprio evento estudado, que também foram retiradas das análises.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Inicialmente, aborda-se a distribuição dos 2062 artigos analisados pelas edições bienais dos eventos, sendo a primeira ocorrida em 1990 e a última em 2013. O Gráfico 1 apresenta estes dados:

Gráfico 1 – Número de artigos de cada edição dos eventos de 1990 a 2013



Fonte: elaborado pelos autores.

É possível identificar um ritmo de crescimento acentuado durante as seis primeiras edições, que correspondem à primeira metade dos eventos analisados. Neste período, o número de artigos aumentou de 22 para 115 da primeira edição (1990) para terceira (1995), representando um aumento de 522,73%. Entre a terceira (1995) e a quinta edição (1999), também houve um aumento significativo no número de artigos, passando de 115 para 196. Embora muito menor que a taxa anterior, este crescimento representa um aumento de 70,43% nos trabalhos apresentados.

O número de artigos, que se conservou constante nas duas edições subsequentes (2001 e 2003), não manteve o ritmo de aumento na segunda metade dos eventos analisados. Embora o número de artigos tenha atingido seu ápice na edição de 2005, com 298 trabalhos, percebe-se uma retração nas edições posteriores. Sendo que se considerada a última edição (2013), houve uma queda de 33,56% no número de trabalhos em relação a 2005. Comparando-se a primeira metade das edições com média de 113,33 artigos (desvio padrão de 69,62) e a segunda, com média de 230,33 (desvio padrão de 35,14) é possível apontar, de maneira geral, um aumento no número de trabalhos apresentados. Além disso, estes dados indicam uma diminuição na variância dos valores apresentados para cada edição, o que pode representar uma estabilização no número de artigos.

Quanto ao processo de mineração de texto, como primeiro produto da aplicação do PreText aos resumos analisados, temos a formação dos *stems*, expressões que representam o núcleo das palavras. Em relação aos termos mais utilizados, e a posição

do termo saúde em relação a estes, segue a análise dos *stems* 1-grama, 2-grama, 3-grama e 4-grama. As Tabelas 1, 2, 3 e 4 apresentam estes *stems* encontrados, a frequência (*fi*) de cada um, e a quantidade (*n*) de resumos em que cada *stem* ocorre para a amostra de 2047 artigos.

Para a análise dos *stem* 1-grama, foram obtidos 12486 *stems*. Destes, foram excluídos os usualmente encontrados em resumos, como ‘estudo’, ‘pesquisa’, ‘artigo’, entre outros. Este procedimento resultou nos seguintes 20 *stems* mais frequentes para 1-grama encontrados, que representam temas adjacentes ou abordagens relacionadas ao acesso solar (Tabela 1).

Tabela 1- 1-grama

<i>stems</i>	<i>fi</i>	<i>n</i>	%
termic	24385	1515	74,01
temperatur	23224	1258	61,46
ambient	21524	1792	87,54
sol	19745	1512	73,86
confort	18392	1754	85,69
urban	14931	1088	53,15
projet	14342	1461	71,37
edifici	13234	1310	64,00
natural	10259	1245	60,82
model	10200	1208	59,01
energ	10050	1238	60,48
iluminaca	8895	812	39,67
intern	8850	1248	60,97
ciudad	8665	1230	60,09
espac	8459	1123	54,86
desempenh	8316	1206	58,92
cal	8151	1103	53,88
edificac	8130	1294	63,21
period	8108	1246	60,87
climat	7724	1211	59,16

Fonte: elaborado pelos autores.

Tabela 1 - 2-grama

<i>stems</i>	<i>fi</i>	<i>n</i>	%
confort_termic	5394	1013	49,49
temperatur_ar	3369	561	27,41
radiaca_sol	3428	761	37,18
eficienc_energet	3085	691	33,76
desempenh_termic	3038	574	28,04
confort_ambiental	2903	874	42,70
ambient_constru	2673	976	47,68
iluminaca_natural	2622	461	22,52
umidad_relat	2530	552	26,97
luz_natural	2522	368	17,98
consum_energ	2513	595	29,07
ventilac_natural	1939	413	20,18
energ_eletr	1610	391	19,10
temperatur_intern	1479	334	16,32
temperatur_med	1386	420	20,52
veloc_vent	1174	320	15,63
ambient_intern	1174	452	22,08
iluminaca_artificial	1134	307	15,00
ar_condicion	1098	337	16,46
carg_termic	1054	313	15,29

Fonte: elaborado pelos autores.

A análise dos termos para 1-grama mais frequentes resultou na identificação de dois grandes grupos. Como era de se esperar, tendo em vista que o evento ENCAC/ELACAC tem como foco principal a área de Conforto Ambiental, no primeiro grupo ocorre a predominância de termos relacionados a conforto térmico (“termic”, “temperatur”), ocorrendo em 74,01% e 61,46% do total de artigos, respectivamente. O termo “sol” (relacionado ao Sol) aparece na quarta posição de *stem* mais frequente. O único termo encontrado entre as 20 primeiras ocorrências, diretamente ligado à adequação da edificação ao bem estar do usuário é “confort”, que ocupa a quinta posição na lista de *stems*. A proximidade de todos estes termos na relação de *stems* mais frequentes pode sugerir que os estudos relacionados ao conforto em edificações priorizam a variável térmica, obtida de forma passiva, por meio da exploração do sol como fonte renovável de energia. Merece destaque a ocorrência do termo “iluminaca” na décima segunda posição, e do termo “natural” na nona posição, possivelmente relacionados à obtenção de iluminação por meio da exploração do acesso solar em edificações. No segundo grande grupo identificado para 1-grama, verifica-se a ocorrência de termos relacionados a questões urbanísticas, (“urban”, “projet”, “edifici”, “ciudad”, “espac”, “edificac”), contabilizando 30% do total dos *stems*

avaliados. Estes termos tratam da morfologia urbana e da geometria das edificações, conceitos dialeticamente envolvidos no processo de favorecimento de condições físicas para o aproveitamento dos benefícios oriundos do acesso solar. Quanto ao *stem* “saud”, relacionado à saúde, este aparece na 643ª posição da listagem para 1-grama (de um total de 12486 posições), ocorrendo 940 vezes em 366 artigos (17,87% da população de artigos avaliada).

Para os *stems* 2-grama, (formados pela ocorrência de dois termos adjacentes), o software PreText 2 formou 65408 *stems*. Na Tabela 2, são apresentadas as 20 primeiras ocorrências encontradas.

De forma similar ao encontrado nos *stems* 1-grama, nos *stems* 20-grama há a predominância de termos relacionados à área de Conforto Ambiental, com dez ocorrências relacionadas ao conforto térmico (“confort_termic”, “temperatura_ar”, “radiaca_sol”, “desempenh_termic”, “confort_ambiental”, “umidad_relata”, “temperatura_intern”, “temperatura_med”, “ar_condicion” e “carg_termic”), representando 50% do total de 20 *stems* avaliados.

O segundo grupo de termos identificado relaciona-se à questão da eficiência energética, com quatro termos identificados (“eficienc_energet”, “consum_energ”, “energ_eletr”, “iluminaca_artificial”), representando 20% do total de *stems*.

Quanto ao tema acesso solar, foram identificados termos relacionados à iluminação natural, representado por três *stems*, representando 15% do total avaliados (“iluminaca_natural”, “luz_natural”, “ambiente_inter”).

Tabela 2 - 3-grama

<i>stems</i>	<i>fi</i>	<i>n</i>	%
confort_ambient_constru	1084	580	28,33
umidad_relata_ar	1015	328	16,02
desempenh_termic_edificac	665	311	15,19
consum_energ_eletr	650	221	10,80
avaliaca_por_ocupaca	535	175	8,55
radiaca_sol_diret	499	244	11,92
clim_quent_umid	487	227	11,09
avaliaca_desempenh_termic	454	205	10,01
nivel_pressa_sonor	452	108	5,28
nivel_eficienc_energet	447	74	3,62
eficienc_energet_edificac	419	223	10,89
radiaca_sol_incident	306	139	6,79
tecnolog_ambient_constru	304	204	9,97
uso_ocupaca_sol	296	137	6,69
sistem_iluminaca_artificial	282	106	5,18
aproveit_luz_natural	277	90	4,40
sistem_condicion_ar	269	91	4,45
incidenc_radiaca_sol	263	166	8,11
habitac_inter_social	257	121	5,91
avaliaca_confort_termic	250	116	5,67

Fonte: elaborado pelos autores.

“incidenc_radiaca_sol”, “avaliaca_confort_termic”), com nove ocorrências, 45% do total de *stems* avaliados.

O segundo grande grupo de *stem* 3-grama identificados são relacionados a conforto ambiental e qualidade do ambiente construído (“confort_ambient_constru”, “avaliaca_por_ocupaca”, “tecnolog_ambient_constru”, “habitac_inter_social”), 20% do total de *stems*. Em terceiro lugar, foram identificados *stems* relacionados a eficiência

Quanto ao tema saúde, o primeiro *stem* encontrado, relaciona-se a saúde pública (“saud_public”) e ocupou a 3996ª posição do total de 65408 *stems* 2-grama obtidos, ocorrendo 55 vezes em 37 artigos (1,80% do total de artigos avaliados).

Na sequência, são apresentados os resultados para a análise 3-grama. Foram formados 18690 *stems* para os artigos da população avaliada; as 20 primeiras ocorrências são apresentadas na Tabela 3.

Confirmando a tendência identificada na análise dos conjuntos de *stems* 1-grama e 2-grama, para 3-grama (18690 ocorrências) prevaleceram os termos relacionados ao conforto térmico (“umidad_relata_ar”, “desempenh_termic_edificac”, “radiaca_sol_diret”, “clim_quent_umid”, “avaliaca_desempenh_termic”, “radiaca_sol_incident”, “sistem_condicion_ar”,

energética (15% do total – “consum_energ_eletr”, “nivel_eficienc_energet”, “eficienc_energet_edificac”). Por fim, temas relacionados a acesso solar e iluminação representam 10% do total (“sistem_iluminaca_artificial”, “aproveit_luz_natural”).

O tema saúde surgiu no *stem* “saud_fisic_mental” de posição 6803 (em um total de 18690), ocorrendo 11 vezes em nove artigos. Isso representa 0,53% do total de artigos avaliados.

Tabela 3 - 4-grama

<i>stems</i>	<i>f_i</i>	<i>n</i>	%
nivel_eficienc_energet_edifici	204	56	2,74
eficienc_energet_edifici_comerci	194	59	2,88
confort_ambiental_eficienc_energet	193	119	5,81
temperatur_umidad_relat_ar	193	114	5,57
tecnic_qualidad_nivel_eficienc	163	53	2,59
qualidad_nivel_eficienc_energet	161	53	2,59
temperatur_ar_umidad_relat	121	72	3,52
confort_termic_espac_abert	110	30	1,47
habitac_unifamili_inter_social	104	87	4,25
diretriz_construt_habitac_unifamili	95	83	4,05
construt_habitac_unifamili_inter	95	83	4,05
bioclimat_brasile_diretriz_construt	94	82	4,01
zoneament_bioclimat_brasile_diretriz	94	82	4,01
avaliaca_desempenh_termic_edificac	92	62	3,03
estrateg_condicion_termic_pass	86	53	2,59
incidenc_radiaca_sol_diret	83	55	2,69
desempenh_termic_edificac_part	76	60	2,93
avaliaca_ruid_are_habit	75	55	2,69
iluminaca_natural_ambient_intern	72	34	1,66
reduca_consum_energ_eletr	68	44	2,15

Fonte: elaborado pelos autores.

da amostra avaliada.

Habitação é o terceiro tema mais recorrente, com três ocorrências, 15% do total (“habitac_unifamili_inter_social”, “diretriz_construt_habitac_unifamili”, “construt_habitac_unifamili_inter”). Já o tema iluminação e acesso solar aparece em quarto lugar, com dois *stems*, 5% do total (Iluminação natural” e “iluminaca_natural_ambient_intern”) e por fim, dois *stems* (10% do total) tratam de assuntos relacionados a conforto ambiental (“bioclimat_brasile_diretriz_construt” e “zoneament_bioclimat_brasile_diretriz”).

O termo saúde “iluminaca_relaca_saud_estar” (aparece na 4929ª posição da listagem para 4-grama (de um total de 4291 posições), ocorrendo seis vezes em cinco artigos (0,24% da população de artigos avaliada).

A partir da análise 5-grama, os resultados para as sequências de *stems* passaram a formar grupos que não atendiam aos critérios de análise adotados, uma vez que não foi encontrada nenhuma sequência de cinco *stems* que tenha ocorrido em pelo menos quatro artigos distintos.

Por fim, são apresentados os resultados obtidos para 4-grama (Tabela 4). Foram obtidos 4291 ocorrências para 4-grama. O tema conforto térmico permanece o mais recorrente entre os *stems* encontrados (“temperatur_umidad_relat_ar”, “temperatur_ar_umidad_relat”, “confort_termic_espac_abert”, “avaliaca_desempenh_termic_edificac”, “estrateg_condicion_termic_pass”, “incidenc_radiaca_sol_diret”, “desempenh_termic_edificac_part”), com sete ocorrências (35% do total).

Em segundo lugar, seis termos relacionados à eficiência energética são identificados “nivel_eficienc_energet_edifici”, “eficienc_energet_edifici_comerci”, “confort_ambiental_eficienc_energet”, “tecnic_qualidad_nivel_eficienc”, “qualidad_nivel_eficienc_energet”, “reduca_consum_energ_eletr”, representando 30% do total de *stems*

5 CONSIDERAÇÕES FINAIS

É necessário considerar que os resultados obtidos estão condicionados aos critérios de seleção e análise adotados; assim, a seleção dos artigos não abarca a totalidade das publicações nacionais existentes relacionadas a acesso solar. Pelo fato do estudo não ser estatístico, mas exploratório, os resultados não podem ser generalizados, devendo ser entendidos unicamente como indicativos. Pelo mesmo motivo, a possibilidade de explicação dos dados sob a ótica de estudos de mesma natureza é improvável. No entanto, na medida em que as evidências encontradas foram apresentadas, poderão ser submetidas a uma reflexão empírica, de forma a prover elementos necessários à elaboração de hipóteses a serem testadas em pesquisas posteriores.

Ressalta-se que a mineração esteve focada neste estudo nos anais do ENCAC/ELACAC desde sua primeira edição, uma vez que se pretendia verificar o interesse de pesquisadores da área de arquitetura e do ambiente construído pelo assunto. Outro enfoque, mais relacionado à área da saúde, poderia ser uma interessante alternativa para embasar a argumentação exposta neste artigo, mostrando as lacunas existentes entre as áreas e, dessa forma, justificar a necessidade da abordagem em futuras pesquisas.

Quanto às limitações do tema, cita-se duas, oriundas da metodologia adotada, que se relacionam à operacionalização da pesquisa: a adoção do número mínimo de artigos para que os *stems* fossem recuperados (4), e a quantidade de *stems* analisados para cada n-grama (20). No entanto, estes parâmetros de análise poderão ser alterados, conforme a necessidade, as condições, e os objetivos de pesquisas futuras. Uma terceira limitação a ser citada é a dificuldade de se pretender vislumbrar determinada tendência tomando por base apenas termos isolados e/ou adjacentes. Nesse sentido, certamente a visualização da totalidade do contexto forneceria mais subsídios. No entanto, dada à inexistência atual de outras técnicas que permitam a análise simultânea da quantidade de artigos avaliada (2047), acredita-se que a forma como as evidências encontradas foram apontadas não comprometa a validade do estudo.

Considerando-se o objetivo da pesquisa, pode-se relatar que quanto ao levantamento sobre a produção científica nacional publicada nos anais do ENCAC/ELACAC relacionada aos termos mais empregados, os mesmo são apresentados nas Tabelas 1, 2 e 3. De forma geral, os resultados obtidos permitiram observar a ocorrência de temas relacionados ao acesso solar localizados três grandes grupos: questões relacionadas ao conforto ambiental (com destaque ao conforto térmico), questões relacionadas ao conceito de eficiência energética e questões relacionadas à iluminação.

Quanto à posição do termo saúde, o mesmo ocorre 940 vezes em 366 artigos para 1-grama (17,87% do total); 55 vezes em 37 artigos (1,80% do total) para 2-grama; 11 vezes em nove artigos para 3-grama (0,53% do total) e seis vezes em cinco artigos para 4-grama (0,24% do total).

Para a população avaliada (2047 artigos) somente 17,87% em algum momento cita o termo saúde, em diferentes contextos. É relevante também a inexistência da ocorrência de termos relacionados às moléstias que podem advir da falta de acesso solar, bem como dos mecanismos biológicos envolvidos em regulação de ciclos naturais (como o circadiano) e na produção hormonal (melatonina e cortisol).

Assim, a produção relativa às relações entre a forma de provimento de acesso ao sol, e quais seriam as condições mínimas para a garantia da ocorrência satisfatória das funções biológicas dos seres humanos permanece pouco explorada, ao se avaliar os resultados da população estudada, trazendo a possibilidade de inferências em relação ao grau de interesse da comunidade

científica acerca do tema, bem como diretrizes para a determinação de temas que carecem de maior estudo, facilitando o alinhamento de pesquisadores da área acerca das possibilidades e dos temas relevantes ainda pouco explorados.

REFERÊNCIAS

- AGGARWAL, C.; ZHAI, C. An Introduction to Text Mining. In: AGGARWAL, C. C.; ZHAI, C. (Eds.). **Mining Text Data**. Springer US, 2012. p. 1–10.
- BAKER, N.; STEEMERS, K. **Daylight Design of Buildings**. Londres: James & James, 2002.
- BOUBEKRI, M. **Daylighting, architecture and health**. Routledge, 2008.
- BUSTOS ROMERO, M. A. **Princípios bioclimáticos para o desenho urbano**. Brasília: CopyMarket, 2000.
- BUTTI, K.; PERLIN, J. **Un hilo dorado**. Madri: Hermann Blume, 1980.
- ESCH, M. M. E. VAN; LOOMAN, R. H. J.; BRUIN-HORDIJK, G. J. DE. The effects of urban and building design parameters on solar access to the urban canyon and the potential for direct passive solar heating strategies. **Energy and Buildings**, v. 47, p. 189–200, 2012.
- FAN, W. et al. Tapping the power of text mining. **Communications of the ACM**, v. 49, n. 9, p. 76–82, 1 set. 2006.
- FARO, A.; GIORDANO, D.; SPAMPINATO, C. Combining literature text mining with microarray data: advances for system biology modeling. **Briefings in Bioinformatics**, v. 13, n. 1, p. 61–82, jan. 2012.
- GAJZLER, M. Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry. **Duomenų Rinkimo Metodai Statybos Sprendimų Paramos Sistemai**, v. 16, n. 2, p. 219–232, jun. 2010.
- HECHT, J. Better than sunshine: see life in an improvement light. **New Scientist**, 06 jul. 2012. Disponível em: <<http://www.newscientist.com/article/mg21428710.300-better-than-sunshine-see-life-in-an-improved-light.html>>. Acesso em: 13 abr. 2014.
- KNOWLES, R. L. The solar envelope: it's meaning for energy and buildings. **Energy and Buildings**. Los Angeles, v.35, n.1, p.15-25, 2003.
- KUMAR, L.; BHATIA, P. K. Text mining: concepts, process and applications. **Journal of Global Research in Computer Science**, v. 4, n. 3, p. 36–39, 2013.
- LABORATORY OF COMPUTATIONAL INTELLIGENCE. **PreText 2: a reestruturação da ferramenta de pré-processamento de textos**. São Carlos: LABIC/ICMC/USP, 2013. Disponível em <<http://sites.labic.icmc.usp.br/pretext2>>. Acesso em: 13 abr. 2014.
- LAMBERTS, R.; DUTRA, L.; PEREIRA, F. O. R. **Eficiência energética na arquitetura**. PW Editores, 1997.
- LITTLEFAIR, P. Daylight, sunlight and solar gain in the urban environment. **Solar Energy**, v. 70, n. 3, p. 177–185, 2001.
- LOBACCARO, G.; FRONTINI, F. Solar Energy in Urban Environment: How Urban Densification Affects Existing Buildings. **Energy Procedia**, v. 48, p. 1559-1569, 2014.
- FONTOYNONT, M. **Daylighting Performance in Buildings**. James and James, London, 1999.
- MARDALJEVIC, J.; CHRISTOFFERSEN, J. A roadmap for upgrading national/EU standards for daylight in buildings. In: CIE Centenary Conference “Towards a new century of light”. Paris: CIE, 15-16, abril. 2013. **Proceedings**. . pp. 178-187.
- MESA, N. A.; CORICA, L.; PATTINI, A. Evaluation of the potential of natural light to illuminate buildings in dense urban environment. A study in Mendoza, Argentina. **Renewable Energy**, v. 36, n. 9, p. 2414-2423, 2011.
- NIE, J.-Y.; GAO, J.; CAO, G. Translingual Mining from Text Data. In: AGGARWAL, C. C.; ZHAI, C. (Eds.). **Mining Text Data SE**. Springer US, 2012. p. 323–359.
- PRADO, A. L. Em busca da pertinência para uma arquitetura tropical. **MDC Revista de Arquitetura e Urbanismo**. Belo Horizonte, Oficina 3 Consultores Associados, n.1, 2 e 3, p. 10-13, 2006. Disponível em: <<http://mdc.arq.br/2006/01/31/em-busca-da-pertinencia-para-uma-arquitetura-tropical/>>. Acesso em: 02 dez. 2013.
- PEDRAZA, E. T.; KUNZE, A.; ROCCASALVA, G.; SCHMITT, G. Best Practices for Urban Densification. In: Computation and Performance - **Proceedings of the 31st eCAADe Conference**, 41-50. Vol. 1. eCAADe: Conferences I. Delft, The Netherlands: Delft University of Technology, 2013.
- RAMIRES, J. C. DE L. O processo de verticalização das cidades brasileiras. **Boletim de Geografia**, v. 16, n. 1, p. 97–106, 2011.
- SEONG, Y.-B. et al. Solar rights analysis system for apartment buildings. **Solar energy**, v. 80, n. 6, p. 723–741, 2006.
- SOARES, M. V. B.; PRATI, R. C.; MONARD, M. C. **Pretext II: Descrição da reestruturação da ferramenta de pré-processamento de textos**. ICMC-USP, 2008.