

食品标签营养素参考值舆情监测、网络关注度与预测

The public opinion monitoring, network attention and forecasting the reference values of the labeled food nutrition

黄泽颖

HUANG Ze-ying

(农业农村部食物与营养发展研究所, 北京 100081)

(Institute of Food and Nutrition Development, Ministry of Agriculture and Rural Affairs, Beijing 100081, China)

摘要:采用 2013—2020 年新浪微博的 1 531 个评论数据和 2 587 个百度指数数据,基于词云分析、文本情感分析以及 ARIMA-SVM 预测模型发现,居民最关注能量的营养素参考值在月饼、酱油、元宵、牛奶等产品使用较多,尤其是儿童食品;总体上,营养素参考值获得略高的好评,但也存在不少争议;营养素参考值没有获得良好的网络关注度,且不随时间同步增长,2020—2022 年,营养素参考值预计受关注的积极性仍不高,亟需科普宣传改进、增添备受关注的营养素,以及推行结合营养素参考值的 FOP 标签,促进营养素参考值发挥应有的效用。

关键词:营养素参考值;食品标签;文本情感分析;网络关注度;ARIMA-SVM 模型

Abstract: Using 1 531 comment data of Sina Weibo and 2 587 of Baidu index data from 2013 to 2020, based on word cloud analysis, text sentiment analysis, and ARIMA-SVM prediction model, based on word cloud analysis, text emotion analysis and ARIMA-SVM prediction model. It was found that residents most concerned about the nutrient reference value of energy, and the nutrient reference value of moon cakes, soy sauce, yuanxiao, milk and other products were more used, especially children's food. In general, the nutrient reference values obtained slightly higher praise, but there were also many controversies; nutrient reference values did not get good attention in the network, and did not increase with time; from 2020 to 2022, the enthusiasm of the nutrient reference values to be concerned will still be not high.

基金项目:中央级公益性科研院所基本科研业务费专项(编号:1610422020006)

作者简介:黄泽颖(1987—),男,农业农村部食物与营养发展研究所副研究员,博士。E-mail:huangzeying@caas.cn

收稿日期:2020-04-09

Therefore, it is urgent to improve the popularization of science, add the nutrients that are concerned, and implement the FOP labeling combined with the nutrient reference values to promote to nutrient reference values play its due role.

Keywords: nutrient reference values; food labeling; text emotion analysis; network attention; ARIMA-SVM model

营养素参考值(Nutrient Reference Values, NRV)作为食品标签营养标示规范化的措施,于 1985 年被国际食品法典委员会(Codex Alimentarius Commission, CAC)提出专门用作食品营养标签标示营养素日需要量的参考值^[1]。营养素参考值表示的是在一份食物中含有的营养成分,可以给人体(特殊人群除外)1 d 提供的营养百分比,例如,100 mL 的牛奶中蛋白质的营养素参考值为 6%,则意味着可以提供给人体 1 d 所需要蛋白质的 6%^[2]。不同国家的营养素参考值根据各国居民的营养素需要量而定,在中国,营养素参考值主要依据中国居民膳食营养素推荐摄入量(RNI)和适宜摄入量(AI)制定。作为中国首个预包装食品营养标签国家标准——《预包装食品营养标签通则》(GB 28050—2011)规定营养成分表由营养成分及其含量、营养素参考值百分比构成,于 2013 年 1 月 1 日起在预包装食品上强制性实施,主要用于标示食品的能量和 4 种核心营养素(蛋白质、脂肪、碳水化合物、钠)。这意味着,营养素参考值以量化的方式在食品外包装袋上标示,能帮助中国居民购买和食用时方便快捷地判断自己摄入的营养素是否足够或超标,以及为选择低热量、低脂肪、低钠的健康食品提供参考依据,对缓和当前中国居民肥胖发生率与营养性疾病增长的凸显矛盾^[3]有重要意义。

营养素参考值已强制实施 8 年,学术界关注主要集

中在营养素参考值标示现状^[4-6]、存在的问题^[7-8]、居民的 KAP(知识、态度、行为)调查^[9-11]等方面。然而,中国居民如何使用与评论,有多大的关注程度,未来关注趋势等问题有待研究。

第 45 次《中国互联网络发展状况统计报告》^[12]指出,截至 2019 年 3 月,中国网民规模达 9.04 亿,即时通信与搜索引擎的用户分别占 99.2% 和 83.0%,从中获取居民的舆情监测和网络关注度等数据全面且客观,具有数据获取渠道的创新性。新浪微博作为中国最大的及时信息分享和传播平台之一,用户的评论数据丰富,而百度是全球最大的中文搜索引擎,其推出的百度指数能客观地反映网民的主动搜索需求和网民对网络信息的关注程度。因此,在大数据时代,文章拟利用互联网大数据开展中国居民对营养素参考值的舆情监测(使用情况与评论)、网络关注度及其趋势分析,对真正发挥营养素参考值的效用,引导食品工业良性发展,保障居民均衡膳食具有重要意义。

1 数据来源与研究方法

1.1 数据来源

文章使用的数据:① 舆情监测数据来自新浪微博上用户对营养素参考值的评论数据。文章以营养素参考值为关键词,网络爬虫收集 2013 年 1 月 1 日—2019 年 11 月 27 日的评论数据 1 531 条。② 网络关注度数据来自百度指数平台 (<http://index.baidu.com>),以“PC 端+移动端”指数(基于中国居民越来越多使用电脑和智能手机搜索信息的趋势)作为搜索指数来源,地区范围设为全国,为提高查全率,以“营养素参考值”为关键词采集 2013 年 1 月 1 日—2020 年 1 月 31 日每日数据,共 2 587 个。

1.2 研究方法

文章运用 R 统计软件,结合营养素参考值数据开展词云分析、文本情感分析与 ARIMA-SVM 模型预测。

1.2.1 词云分析 词云是一种特殊的文本可视化,频率更高的词通过占据突出的位置有效地表示出来^[13]。文章抽取营养素参考值评论文本的关键词进行词频统计,然后利用 R 软件的 Wordcloud 包绘制词频云图,且根据 Wordcloud 函数自行调节关键词、词频、字号、颜色、角度、词频量和关键词个数,字体越大代表频率越高,颜色相同代表频率相同。

1.2.2 文本情感分析 文本情感分析又称情感挖掘、倾向性分析,是对带有情感色彩(如赞成、反对、表扬、批评、喜、怒、哀、乐等)的评论文本内容进行处理、归纳和推理的过程,从而发现评论者对研究对象的态度和意见^[14]。文章通过对营养素参考值的评论文本进行拆句,从中抽取情感词,匹配情感词典,分类合并每条评论 ID,对每条评论文本进行情感倾向性分析,最终根据划分规则,将情感

值 >0 表示为正面情感,情感值 $=0$ 表示为中性情感,情感值 <0 表示为负面情感,从而计算正面情感词、中性情感词、负面情感词的个数,获取整体情感倾向。

1.2.3 ARIMA-SVM 模型预测 自回归积分滑动平均模型 (Autoregressive Integrated Moving Average Model, ARIMA) 是一种时间序列预测方法, 但只适用于解决线性问题, 在处理非完全线性关系时将会出现偏差^[15]。支持向量机 (Support Vector Machine, SVM) 是 Vapnik 等^[16]根据统计学理论提出的机器学习方法, 通过结构风险最小化原则, 较好地解决了小样本、非线性等实际问题。由于营养素参考值的网络关注度受到居民的个人与家庭情况、社会宣传、突发事件等多因素的制约, 且因素间保持着错综复杂的联系, 导致网络关注度具有非平稳性、非线性等复杂的非线性动力系统特征, 单独使用某一种方法进行预测可能导致误差过大。因此, 文章采用 ARIMA-SVM 模型对网络关注度进行预测, 首先通过 2013 年 1 月 1 日—2018 年 12 月 31 日的百度指数建立 ARIMA 模型、SVM 模型、ARIMA-SVM 模型预测 2019 年 1 月—2020 年 1 月每个月份首日的百度指数, 其中, ARIMA-SVM 模型的预测结果是利用原值序列减去 ARIMA 预测序列得到残差序列, 对残差序列 $\{e_t\}$ 使用 SVM 模型进行预测, 高斯核函数和参数采用交叉验证方法进行优化选择, 使用最优参数下训练得到的模型对 e_t 进行预测, 最后根据公式 $\hat{y}_t = \hat{l}_t + \hat{e}_t$ 获得网络关注度预测值。然后, 采用平均绝对百分误差 (MAPE)、均方根误差 (RMSE) 和希尔不等系数 (Theil IC) 3 个指标考察 3 个模型的预测效果, 一般情况下, MAPE 值低于 10、RMSE 值越小、Theil IC 介于 0 与 1 之间且越接近于 0, 则表明预测精度越高。最后, 利用预测效果最优的模型预测 2020—2022 年的百度指数。

2 结果与分析

2.1 词云分析

营养素参考值词云图(图 1)中,要求被标识的“4+1”营养成分中,出现频率从高到低依次是能量、蛋白质、脂

图 1 营养素参考值词云图^[17]

Figure 1 Word cloud of nutrient reference values

肪、维生素、碳水化合物,可见,中国居民在选购预包装食品时,主要关注食品的能量占每天推荐摄入量的比重;食品方面,月饼、酱油、元宵、牛奶的营养素参考值被提及的次数较多,而且,营养素参考值被多数人用于选购儿童食品。

2.2 文本情感分析

中国居民对营养素参考值分别持有中性、负面和正面情感(见表 1),在 1 531 个评论中,正面情感和中性情感的比重约 35%,正面情感略微高,但持负面情感的评论数比重不低(接近 28%),居民较多认为营养素参考值难懂,不易理解含义,没有起到判断食品健康与否的效果,可以判断,居民对营养素参考值的态度是复杂的,既有一定程度的好评,但也存在不好的评价。

表 1 营养素参考值文本情感分析结果^[17]

Table 1 Text emotion analysis results of nutrient reference values

情感	评论数	评分占比/%
正面情感	565	36.90
中性情感	547	35.73
负面情感	419	27.37
总计	1 531	100.00

2.3 营养素参考值网络关注度

中国对营养素参考值的网络关注度见图 2。8 年来中国每日百度指数基本在 200 上下浮动,日均为 232,最低为 12,最高达到 993(2017 年 8 月 30 日),可见,2013 年 1 月 1 日起中国实施《预包装食品营养标签通则》以来,居民持续在关注营养素参考值,但关注度不高,也没有随着国民经济水平、教育水平、健康意识的提高而同步增长,可见,中国实施的营养参考值没有引起居民较大的关注。

中国居民对营养素参考值网络关注度有性别差异,男性占 43.68%,而女性占 56.32%。同样,也存在年龄分布差异(见图 3),20~29 岁的比重最高,接近 44%,其次是 19 岁及以下和 30~39 岁(均占比 25.29%),但 40~49 岁的居民比重不到 5%,特别是 50 岁及以上(1.15%)。

省份方面,中国居民对营养素参考值的日均网络关注度存在差异性(见图 4),广东省最高,超过 100,其次是北京(95)、江苏(95)、浙江(85)、山东(79)、上海(78),总体上是来自发达省市的居民比较关注营养素参考值。

2.4 ARIMA-SVM 模型预测

由表 2 的 $\hat{y}_t$ 可知,ARIMA-SVM 模型的预测结果与实际值 y_t 在 2019 年 1 月、2 月、6 月的首日预测值最为接近。

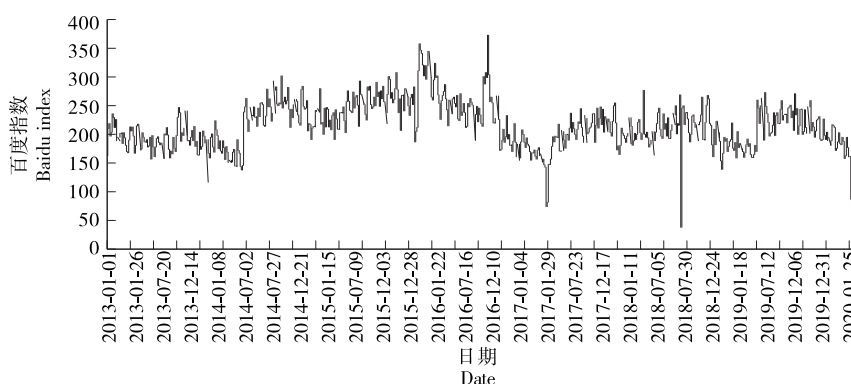


图 2 2013 年 1 月 1 日—2020 年 1 月 31 日中国的营养素参考值百度指数变化趋势

Figure 2 The trend change of the baidu index of nutrient reference values in China from January 1, 2013 to January 31, 2020

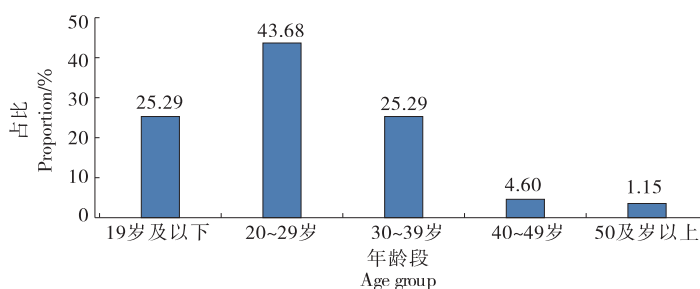


图 3 营养素参考值网络关注度的年龄分布图

Figure 3 The residents' age distribution of network attention of nutrition reference values

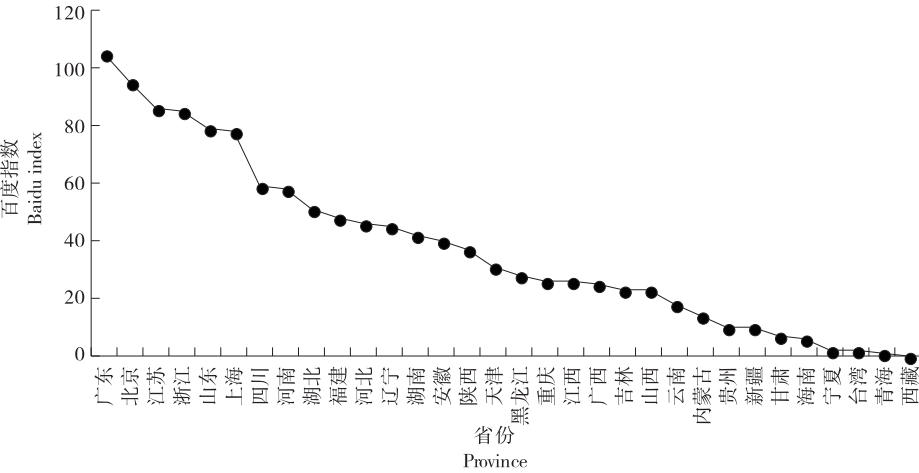


图 4 2013 年 1 月 1 日—2020 年 1 月 31 日日均各省营养素参考值百度指数
Figure 4 Dailybaidu index of nutrient reference values in each province
from January 1, 2013 to January 31, 2020

由 ARIMA 模型、SVM 模型和 ARIMA-SVM 模型对 2019 年 1 月—2020 年 1 月每月首日的预测值与实际值的比较(见表 3)发现,ARIMA-SVM 模型的预测接近程度不如其他两个模型的时间,仅是 2019 年 4 月、11 月的首日,而其他时间的预测效果不差。在一定程度上,ARIMA-SVM 模型的预测效果较佳。

通过 3 个指标的比较(见表 4),ARIMA-SVM 模型比单个 ARIMA 模型与 SVM 模型的预测精度要高,故采用 ARIMA-SVM 模型对营养素参考值网络关注度进行预测。

基于 ARIMA-SVM 模型对 2020 年 2 月—2022 年 12 月每月首日营养素参考值百度指数的预测发现(见图 5),未来3年,中国居民对营养素参考值的网络关注度仍比

较低迷,百度指数在 144~179 范围波动,平均值为 162,比过去 8 年的均值(232)低。

表 3 百度指数实际值与 3 个模型预测值的比较^[18]
Table 3 The comparison between the actual value of baidu index and the predicted value of 3 models

时间	实际值	ARIMA 模型预测值	SVM 模型预测值	ARIMA-SVM 模型预测值
2019-01-01	167	194	190	167
2019-02-01	155	186	182	153
2019-03-01	178	299	179	185
2019-04-01	208	277	176	120
2019-05-01	219	233	173	178
2019-06-01	223	283	170	234
2019-07-01	260	256	167	200
2019-08-01	238	213	164	205
2019-09-01	189	235	161	148
2019-10-01	169	280	158	188
2019-11-01	200	248	155	122
2019-12-01	229	242	152	165
2020-01-01	191	185	150	171

表 4 3 种模型预测效果的比较^[18]
Table 4 Comparison of prediction effects of 3 models

预测模型	MAPE	RMSE	Theil IC
ARIMA 模型	8.708 6	9.129 2	0.185 4
SVM 模型	6.305 4	8.714 4	0.034 7
ARIMA-SVM 模型	2.203 4	4.673 4	0.032 7

时间	实际值 y_t	$\hat{y}_t$	e_t	$\hat{e}_t$	$\hat{y}_t$
2019-01-01	167	173	-6	-6.36	167
2019-02-01	155	161	-6	-7.99	153
2019-03-01	178	167	11	17.57	185
2019-04-01	208	117	91	3.31	120
2019-05-01	219	169	50	9.42	178
2019-06-01	223	213	10	20.85	234
2019-07-01	260	181	79	18.99	200
2019-08-01	238	199	39	6.44	205
2019-09-01	189	125	64	23.34	148
2019-10-01	169	158	11	30.40	188
2019-11-01	200	117	83	5.27	122
2019-12-01	229	149	80	15.89	165
2020-01-01	191	166	25	4.95	171

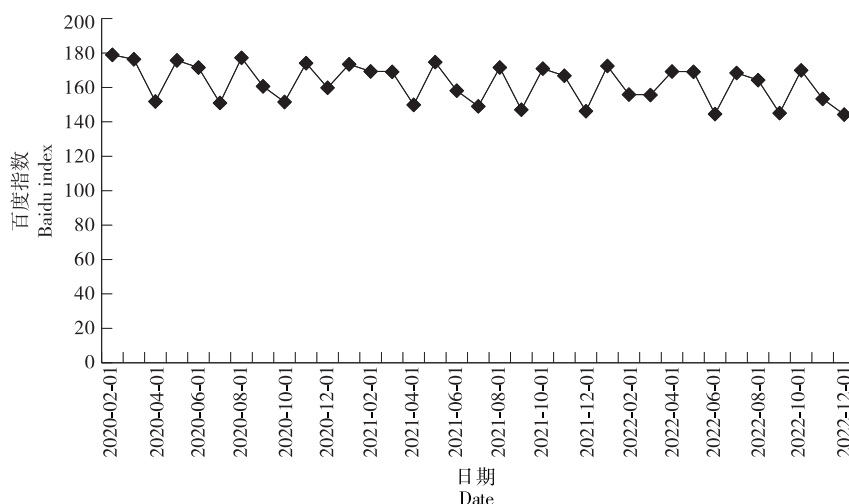


图 5 2020 年 2 月—2022 年 12 月每月首日的营养素参考值百度指数的预测值

Figure 5 The predicted value of baidu index of nutrient reference value on the first day of each month from February 2020 to December 2022

3 结论与建议

为了解食品标签营养素参考值实施 8 年来中国居民的使用情况、评论、网络关注度,以及预测未来 3 年网络关注度的变化情况,文章采用 2013 年以来的新浪微博评论数据和百度指数,基于词云分析、文本情感分析以及 ARIMA-SVM 预测模型,得出的研究结论是,营养素参考值在中国实施没有获得广泛的好评和关注,居民提及最多的是能量、月饼、酱油、儿童食品的营养素参考值,然而表示出复杂的情感,虽然有较高的好评,但也持有不满意的态度;8 年以来,中国居民对营养素参考值的网络关注度平平,百度指数不随时间显著增长;营养素参考值网络关注度存在性别、年龄与省份差异,以女性、20~29 岁、来自发达省市的群体为主;基于预测精准度高的 ARIMA-SVM 模型,2020—2022 年营养素参考值的网络关注度预计仍较低迷。营养素参考值作为国际上常见的食品标签,在中国推行的情况不够乐观。在健康中国建设时期,要使营养素参考值发挥显著的作用,需要一些促进营养素参考值发挥应有作用的配套措施,拟提出如下几点政策建议。

3.1 改进科普宣传,提高使用积极性

要使营养素参考值产生最大的影响,科普宣传是关键^[19]。在中国,健康饮食看营养标签的科普宣传开展不少,主要围绕如何看懂营养素参考值主题,大概介绍了营养素参考值的概念、用途及使用方法。这样的科普固然必要,但断章取义式的科普方式缺少介绍营养素参考值的科学依据、推行的原因以及如何与日常饮食有效结合,使多数人只知其一不知其二,无法真正理解营养素参考值,更别说科学指导日常饮食。

因此,改进营养素参考值的科普宣传极为重要,建议在科普类文章或者常识解说中,以讲故事或举身边老百姓的例子,少用学术用语,改用通俗说法,少用长篇大论,改用诙谐有趣的卡通动漫,少用文字,改用声频与视频,达到简洁明了阐述营养素参考值来龙去脉的目的。此外,建议借鉴美国 1990 年发布的营养标签和教育法(Nutrition Labeling and Education Act, NLEA),确立中国的营养标签教育法,为营养素参考值的科普宣传提供法律依据。

3.2 增添备受关注的营养成分,提高利用率

能量+4 种核心营养素是包装食品强制性标示的营养成分,比较常见,但其他营养成分如膳食纤维、优质蛋白、维生素、矿物质等由于非强制要求,仅在个别食品显示。为满足消费者对更多营养成分的关注需求,提高营养素参考值的利用率,支撑更多食品的健康选购以及适应更广泛的人群,建议修改《预包装食品营养标签通则》,添加糖(单糖与双糖)、膳食纤维、饱和脂肪、反式脂肪等营养成分,显示其营养素参考值,为消费者选择合适、健康的食品提供更多的判断依据。

3.3 尝试推行与营养素参考值相结合的 FOP 标签

片面显示单个或数个营养成分的营养素参考值难以适应大健康时代需求,食品包装正面(Front of Package, FOP)标签通过营养素度量法(Nutrient Profile, NP)向消费者直观易懂地评价食品整体营养价值,克服了营养素参考值认知度低且位置不醒目等问题。一些发达国家在 FOP 标签中融入营养素参考值,例如英国的交通灯信号标签显示营养素参考值,当某一种营养物质的营养素参考值超过 30%(食品)及超过 15%(饮料)时,则会显示红色,警告消费者慎重选购^[20]。在中国,《健康中国行动

(2019—2030年)》已将推动 FOP 标签列入国家行动计划,在改进标签设计与落地实施之前,建议总结英国等国的经验做法与存在的不足,尝试将 FOP 标签与营养素参考值相结合,根据国民营养健康状况与发展目标,对重要的营养素含量根据营养素参考值划分健康等级,并采用颜色编码、图形、文字、数字其中的一种方法标示。

参考文献

- [1] 王杰, 蒯士安. 营养素参考值在食品营养标示中的应用[J]. 中华预防医学杂志, 2006, 40(1): 62-64.
- [2] 曹娅, 邹兰兰. 营养素参考值在食品营养标示中的作用分析[J]. 食品安全导刊, 2019(4): 178-179.
- [3] 国家卫生计生委疾病预防控制局. 中国居民营养与慢性病状况报告: 2015年[R]. 北京: 人民卫生出版社, 2015.
- [4] 刘抚, 陶晔璇, 蔡威. 上海市预包装食品营养标签现状调查[J]. 上海交通大学学报(农业科学版), 2014, 32(5): 70-75.
- [5] 钞凤, 张书芳, 叶冰, 等. 河南省预包装食品营养标签标示现状调查[J]. 中国食品卫生杂志, 2015, 27(4): 437-440.
- [6] 姜迎, 于连龙, 孙青, 等. 7 大类 623 种预包装食品营养标签标示情况调查分析[J]. 中国健康教育, 2018, 34(1): 46-50.
- [7] 刘璐, 谷瑞丽, 刘智勇, 等. 铁路站车销售预包装食品标签问题调查及解析[J]. 食品安全导刊, 2019(12): 96-97, 99.
- [8] 容慧, 李小丽, 罗柳慈, 等. 进口预包装食品标签存在问题及措施研究[J]. 食品科技, 2018, 43(2): 325-330.
- [9] HE Yao, HUANG Li-ping, YAN Si-jin, et al. Awareness, understanding and use of sodium information labelled on pre-packaged food in Beijing: A cross-sectional study[J]. BMC Public Health, 2018(18): 509.
- [10] 彭程煜, 袁桂, 饶春平. 高职学生食品营养标签知识、态度、行为调查[J]. 卫生职业教育, 2017, 35(11): 106-107.
- [11] 马丽萍, 陈琼, 尹逸. 广东省某高职院校食品专业学生关于营养标签的 KAP 调查[J]. 中国食物与营养, 2016, 22(8): 78-81.
- [12] 中国互联网络信息中心. 第 45 次《中国互联网络发展状况统计报告》[EB/OL]. (2020-04-28) [2020-05-01]. <http://cnnic.cn/hlwfzyj/hlwxyzbg/hlwjtjbg/202004/P020200428399188064169.pdf>.
- [13] MCNAUGHT C, LAM P. Using wordle as a supplementary research tool[J]. The Qualitative Report, 2010, 15(3): 630-643.
- [14] PANG Bo, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques[J]. Empirical Methods in Natural Language Processing, 2002 (5): 79-86.
- [15] NEWBOLD P. ARIMA model building and the time series analysis approach to forecasting[J]. Journal of Forecasting, 1983, 2(1): 23-35.
- [16] VAPNIK V. The nature of Statistical learning theory[M]. New York: Springer, 1995: 133-137.
- [17] 新浪网技术(中国)有限公司. 关于营养素参考值的新浪微博评论数据 [EB/OL]. (2019-12-27) [2019-12-28]. <https://s.weibo.com/weibo/%25E8%2590%25A5%25E5%2585%25BB%25E7%25B4%25A0%25E5%258F%2582%25E8%2580%2583%25E5%2580%25BC?topnav=1&wvr=6&b=1>.
- [18] 百度控股有限公司. 关于营养素参考值的百度指数[EB/OL]. (2020-02-01) [2020-02-05]. <http://index.baidu.com/v2/main/index.html#/trend/%E8%90%A5%E5%85%BB%E7%B4%A0%E5%8F%82%E8%80%83%E5%80%BC?words=%E8%90%A5%E5%85%BB%E7%B4%A0%E5%8F%82%E8%80%83%E5%80%BC>.
- [19] BYRD-BREDBENNER C, ALFIERI L, WONG A, et al. The inherent educational qualities of nutrition labels[J]. Family & Consumer Sciences Research Journal, 2001, 29 (3): 265-280.
- [20] UK Department of Health. Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets [EB/OL]. (2016-10-01) [2019-10-15]. <https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance>, 2016.

信息窗

泰国优化了食品备案在线提交系统

泰国食品药品监督管理局 9 月 10 日消息:该部门下属机构食品局发布新闻,公布自 2020 年 9 月 21 日起,部分加工食品可以通过在线新系统提交备案申请。

相关内容如下所述:

(1) 部分加工食品包括:密封包装的饮料、速溶混合咖啡、混合咖啡、速溶咖啡、电解质饮料、牛乳、调味牛乳、其他乳制品、发酵乳、冰淇淋、密封包装的食品。

(2) 新系统相比旧系统的特点:更加方便(系统可以识别申请项目而进行分类,并且仅填写与所提交产

品相关的信息即可,包括要审核的证明文件);更快捷(更快的处理器);减少出现问题和错误的情况。

(3) 对于在旧系统中 2020 年 9 月 21 日之前提交的申请:如果是处在审核中的状态,需要在原系统中跟进并完成申请过程;如果是待付款的状态,需要全部删除(撤回),包括待付款的付款单,以便允许申请者通过新系统进行提交申请。

(来源:<http://news.foodmate.net>)