

Cross-lingual Search Engine for Retrieval of Greenhouse Gas Emission Factors

Nattapot Nuwong¹, Yachai Limpiyakorn²

^{1,2}Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand

Email address: ¹6770233221@student.chula.ac.th, ²yachai.l@chula.ac.th

Abstract— Climate change is a critical global challenge primarily driven by greenhouse gas (GHG) emissions. Managing GHG data effectively is essential for carbon footprint assessments. However, determining the emission factor associated with an activity is not straightforward and labor-intensive. Moreover, the contents in the emission factor database is the mixing of Thai and English languages. This study therefore develops a cross-lingual search engine for retrieval of emission factors from Thailand Greenhouse Gas Management Organization databases to promote the reliability of carbon emission computation. The system enables seamless bilingual search functionality, integrating dictionary-based synonym matching and Elasticsearch indexing. The semantic search with synonym-based search expansion would ensure relevant results regardless of query language. By implementing an automated pipeline for data retrieval and indexing, the system enhances retrieval accuracy and efficiency thereby aiding organizations in precise carbon footprint calculations.

Keywords— Carbon emission: cross-lingual information retrieval: emission factor: semantic search: synonym.

I. INTRODUCTION

The world is currently facing extreme weather conditions, primarily due to the continuous and substantial emissions of greenhouse gases (GHGs) into the atmosphere. These emissions have widespread impacts on ecosystems, climate, biodiversity, and the quality of life at both local and global levels. Consequently, the effective monitoring and management of GHG emissions data is of paramount importance. In Thailand, the principal agency responsible for assessing, collecting, compiling, and disseminating GHG-related information is the Thailand Greenhouse Gas Management Organization (TGO), which operates under the Ministry of Natural Resources and Environment. TGO has developed comprehensive databases and reports covering a wide range of sectors including industry, energy, agriculture, and communities as well as information for assessing and calculating carbon footprints. Moreover, it publicizes greenhouse gas emissions factors based on Thai National LCI Database [1], in addition to international standards and guidelines such as Intergovernmental Panel on Climate Change (IPCC) [2], and ecoinvent [3]. Currently, these emission factors are published on the TGO website [4] and users can download the electronic files. It remains a gap the absence of a search system that can comprehensively handle emission factor queries in a cross-lingual context. In the content provided on the TGO Website, some EF keywords appear in English, some in Thai, and some in both languages. Additionally, user preferences vary, i.e. some users favoring

Thai search terms while others prefer English. These linguistic discrepancies can cause conventional single-language search approaches to yield inaccurate or incomplete result. This research proposes the development of a cross-lingual search engine for retrieving GHG emission factors from TGO databases. The system is designed to support semantic search, incorporating synonym matching to enhance user accessibility regardless of language.

II. BACKGROUND

A. Greenhouse Gas Emission Factor

The greenhouse gas emission factor is calculated based on the amount of greenhouse gas emitted and absorbed per unit of activity. It is used to assess the volume of greenhouse gas emission resulting from various activities, such as energy consumption, fuel consumption, transportation, or industrial production processes. This factor is a critical variable that enables organizations, government agencies, small businesses to calculate and analyze carbon emissions. It is essential for reporting, impact monitoring, and planning strategies to reduce greenhouse effects.

Referring to the Kyoto Protocol, the recognized GHGs include Carbon Dioxide (CO₂), Methane (CH₄), Nitrous Oxide (N₂O), Hydrofluorocarbons (HFCs), Perfluorocarbons (PFCs), Sulphur Hexafluoride (SF₆), and Nitrogen Trifluoride (NF₃). Greenhouse gas emissions are expressed in terms of kilograms of carbon dioxide equivalent (kgCO₂e), a standard metric referred to as carbon emissions.

The formula for calculating carbon emissions is as follow.

$$\text{Carbon Emissions} = \text{Activity Data} \times \text{Emission Factor}$$

where Emission Factor (EF) is a coefficient that describes the rate at which a given activity releases greenhouse gases into the atmosphere.

B. Carbon Footprint

The carbon footprint assesses the overall greenhouse gas emissions—both direct and indirect—related to a service, product, or activity. The calculation of carbon footprint can be separated into two types:

- Carbon Footprint for Organization (CFO)
- Carbon Footprint of Product (CFP)

CFO evaluates the total greenhouse gas emissions released by an organization over a specific period, saying annually. It comprises direct emissions (Scope 1), indirect emissions from energy consumption (Scope 2), and additional indirect emissions (Scope 3). While CFP assesses the total greenhouse

gas emissions released throughout the life cycle of a product consisting five main stages:

- raw material procurement
- production
- transportation
- use
- waste disposal

C. Cross-Lingual Information Retrieval (CLIR)

CLIR is a specialized area of information retrieval that deals with retrieving relevant information from a document collection written in a language different from the user's query language. It addresses the challenge of matching queries and documents across different linguistic boundaries. This technology has become increasingly important in our globalized information landscape where valuable content exists across linguistic boundaries. In literature, the authors [5] developed a CLIR system designed to enhance resource accessibility for English language learners from diverse linguistic backgrounds. The proposed method outperformed all others with an exceptional accuracy of 98.82%. The study [6] reported that bilingual semantic networks enable effective query translation and expansion in cross-lingual information retrieval. By mapping equivalent concepts across languages, these networks allow users to retrieve relevant documents in a target language even when queries are formulated in a different language. This approach overcomes limitations of simple dictionary-based translation by capturing the complex semantic relationships specific to specialized domains.

D. Synonyms

Although they have distinct forms, synonyms are words that share similar or identical meanings and can be used interchangeably in some situations [7]. Synonyms add variety to a language, less repetitive, and enable more precise and nuanced communication.

Synonym classification by linguistic origin [8, 9] :

- Monolingual Synonyms: Words that originate from the same language, such as "freedom" and "liberty" in English.
- Cross-linguistic Synonyms: Words that come from different languages but have the same meaning, such as "โรงหนัง" (Thai) and "movie theater" (English).

E. Hybrid Search

Hybrid search combines multiple search paradigms to deliver more relevant results than any single approach could achieve alone [7]. The concept of hybrid search emerges from the recognition that traditional search methodologies such as keyword-based lexical search and dense vector search, can be considered orthogonal approaches with distinct advantages, offering complementary strengths. Lexical search excels at precision for exact matches, while vector search captures semantic relationships that might not share exact terminology.

III. METHODOLOGY

The research employs a cross-lingual information retrieval approach combining Elasticsearch with Thai-English synonyms dictionary for enabling seamless search across both

languages. Visit the created synonyms dictionary at <https://tinyurl.com/ef-dict>. The search engine returns emission factor data from the database constructed from the TGO databases. Excerpts of the database created in this work is shown in Table 1.

Fig. 1 depicts the methodology presented in this study. The details are elaborated in the following sections.

A. Create Synonyms Dictionary for CLIR

The synonyms generation process aims to identify and analyze pertinent terminology for individual entries residing TGO databases. The method commences by incorporating each item with synonyms derives from keywords extracted from its description, synonyms suggested by theecoinvent databases, or lexical references to facilitate translation and semantic alignment between Thai and English terminologies. Another source of synonyms comes from vernacular expressions such as ก๊าซหุงต้ม, meaning cooking gas, is paired with Liquefied Petroleum Gas or LPG. Furthermore, the process assists in discovering additional lexical items, exemplified by the correlation between "gasoline" and "E85" within the context of automotive fuel.

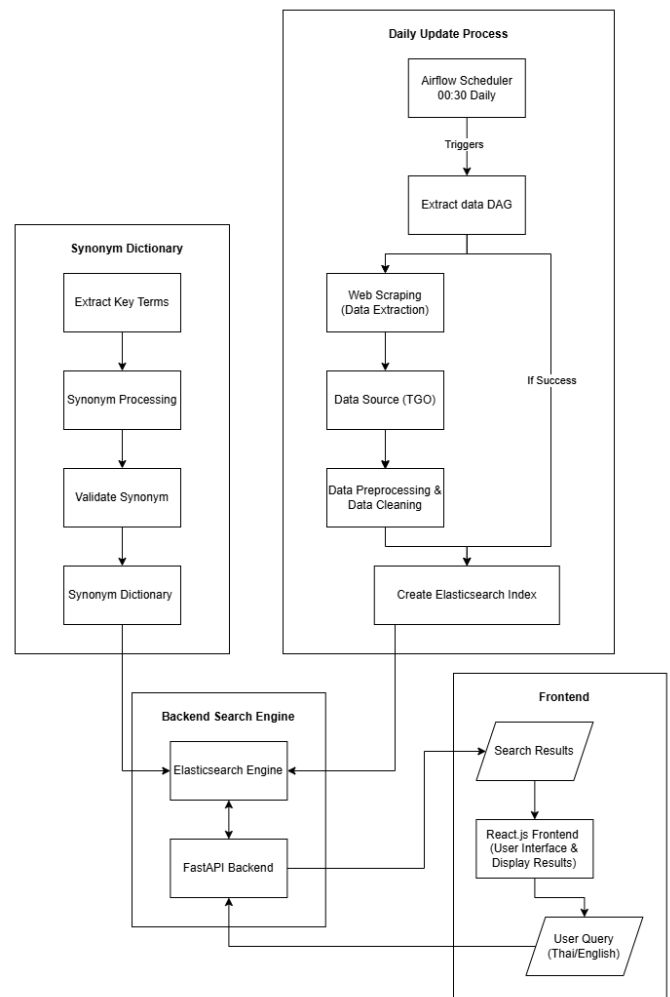


Fig. 1. Research methodology.

TABLE I. Excerpt of constructed emission factor database.

Category	Name	Unit	Factor (kgCO ₂ e/Unit)	Reference	Last Updated
Electricity	Electricity, grid mix (ไฟฟ้า)	kWh	0.5986	Thai National LCI Database, TIIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019
Tap water	น้ำประปา-การประปานครหลวง	m3	3.2281	Thai National LCI Database, TIIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019
Natural gas product	ก๊าซธรรมชาติแบบผสม	kg	0.7544	Thai National LCI Database, TIIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019
Natural gas product	ก๊าซธรรมชาติแบบผสม	m3	0.5767	Thai National LCI Database, TIIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019
Liquid fuels and solid fuels	Gasoline (แก๊สโซลีน)	kg	0.4024	Thai National LCI Database, TIIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019
Liquid fuels and solid fuels	Diesel (น้ำมันดีเซล น้ำมันโซลาร์)	kg	0.3522	Thai National LCI Database, TIIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019
Liquid fuels and solid fuels	ซัลเฟอร์	kg	0.2390	Thai National LCI Database, TIIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019
Chemicals	Sulfur	kg	0.4529	Thai National LCI Database/MTEC	24Sep 2012
Agriculture goods and food	กาแฟอาราบิก้า	kg	11.0291	Thai National LCI Database, TIIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Mar 2021
Pulp and Paper	กระดาษพิมพ์เขียนแบบไม่เคลือบผิว	kg	2.1020	Thai National LCI Database, TIIS-MTEC-NSTDA	Dec 2019
Textiles	คาร์บอนไฟเบอร์	kg	7.5500	Ecoinvent 2.0 (ส่วนผสมของ graphite และในล่อน)	Aug 2011
Paper and Packaging	Supreme A4 White Copy Paper 80gsm 1 sheet	sheet	0.0115	PHOENIX PULP & PAPER PUBLIC COMPANY LIMITED	Nov 2012
Petrochemical	Soaping agent	kg	2.3372	Ecoinvent 2.2, IPCC 2007 GWP 100a	24Sep 2012

B. Establish Emission Factor Database

The emission factor database has been created and updated by collecting the data from several TGO websites, including

- Emission Factor (CFO) Scope 1, 2 [10]
- Emission Factor (CFP)/ (CFO) Scope 3 [11]
- Carbon Label Project [12]. The carbon label notifies the total greenhouse gas emissions released by an individual product brand and type.

The created emission factor database contains totally 9,990 records that collected from emission factor CFO database 49 records, emission factor CFP database 762 records, and Carbon Label database 9,179 records.

C. Analyzers in Elasticsearch

In Elasticsearch [13], configuring index settings utilizes multiple analyzers to effectively process natural language text and improve search accuracy [14]. The configuration settings are shown in Fig. 2, including

Filter Configurations setup:

- `thai_ngram_filter`: The N-Gram filter is specifically configured for Thai language text analysis since the language naturally lacks space-based word boundaries.
- `my_stop_filter`: The stop words filter is designed to remove common words that do not carry significant meaning, such as "the" and "with" in English.
- `thai_english_synonym_filter`: The filter references a comprehensive synonym dictionary explicitly defines equivalent terms across Thai and English languages.

Main Analyzer Definitions:

- `thai_synonym`: An analyzer uses "icu_tokenizer" or 'thai' suitable for tokenizing Asian languages including Thai, and incorporates the dictionary-based synonym resolution "thai_english_synonym_filter" to map terms to their semantic equivalents.

- `thai_synonym_analyzer`: Another specialized analyzer for explicit synonym mapping using a standard tokenizer and applying the dictionary-based "thai_english_synonym_filter" to ensure consistent term normalization.

IV. DEMONSTRATION

The implemented search functionality using *Hybrid search* with the configured *synonym* dictionary was evaluated using several case studies across various domains.

```

index_settings = {
  "settings": {
    "analysis": {
      "filter": {
        "thai_edge_ngram_filter": {
          "type": "edge_ngram",
          "min_gram": 1,
          "max_gram": 20,
          "token_chars": ["letter", "digit", "whitespace"]
        },
        "thai_english_synonym_filter": {
          "type": "synonym_graph",
          "synonyms_path": "analysis/synonyms.txt",
          "expand": true,
          "updateable": true
        }
      },
      "analyzer": {
        "thai_autocomplete_analyzer": {
          "type": "custom",
          "tokenizer": "icu_tokenizer",
          "filter": [
            "lowercase",
            "icu_folding",
            "thai_edge_ngram_filter"
          ]
        },
        "thai_synonym_analyzer": {
          "type": "custom",
          "tokenizer": "icu_tokenizer",
          "filter": [
            "lowercase",
            "icu_folding",
            "thai_english_synonym_filter"
          ]
        }
      }
    }
  }
}

```

Fig. 2. Configuration Settings in Elasticsearch.

A. Case Study 1: “ก๊าซหุงต้ม”

“ก๊าซหุงต้ม” can be translated to English, namely “cooking gas”, “Liquefied Petroleum Gas” and “LPG”. In CFO Scope 1 dataset, the key search “ก๊าซหุงต้ม” matches “LPG”(litre) and “LPG”(kg) categorized in Stationary Combustion, in addition to matching “Liquefied Petroleum Gas”(litre), and “Liquefied Petroleum Gas”(kg) categorized in Mobile Combustion (on road). Further, “ก๊าซหุงต้ม” can be considered a product of two categories: ก๊าซธรรมชาติ or Natural gas (natural gas separation process), and พลังงาน or Fuel (crude oil refinery process). The excerpt of search results is shown in Fig. 3. The total search results found 50 entries of which most entries match “แอลพีจี” (LPG) residing Carbon Label Project dataset. Some match “Gasoline (แก๊สโซลีน)” categorized in Fuel of CFP dataset and Petroleum in Carbon Label Project dataset.

50 results found for “ก๊าซหุงต้ม”
50 Activities, 50 Emission factors

Showing items 1 - 15 of 50 total

Category	Name	Unit	Factor (kgCO ₂ e/Unit)	Reference	Last Updated	Score
กลุ่มพลังงาน: เชื้อเพลิงเหลวและเชื้อเพลิงแข็ง	Liquefied Petroleum Gas, LPG Mixed (ก๊าซหุงต้มแบบผสม)	kg	0.8582	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019	102.33402
กลุ่มผลิตภัณฑ์จากก๊าซธรรมชาติ	ก๊าซหุงต้ม	kg	1.1839	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019	57.13681
กลุ่มพลังงาน: เชื้อเพลิงเหลวและเชื้อเพลิงแข็ง	ก๊าซหุงต้ม	kg	0.4267	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019	57.13681
Stationary Combustion	LPG	litre	1.68122603	IPCC Vol.2 table 2.2, DEDE	Apr 2022	48.23645
Stationary Combustion	LPG	kg	3.113381537	IPCC Vol.2 table 2.2, DEDE LPG 1 litre = 0.54 kg	Apr 2022	48.23645
Mobile Combustion (On road)	Liquefied Petroleum Gas	litre	1.73064606	IPCC Vol.2 table 3.2.1, 3.2.2, DEDE	Apr 2022	43.58485
Mobile Combustion (On road)	Liquefied Petroleum Gas	kg	3.204900111	IPCC Vol.2 table 3.2.1, 3.2.2, DEDE LPG 1 litre = 0.54 kg	Apr 2022	43.58485

Fig. 3. Search results of “ก๊าซหุงต้ม”.

B. Case Study 2: “กาแฟ Arabica”

In Thailand, Office of the Royal Society is in charge of transliteration from English to Thai. For example, “อะราบิก้า, อามาริก้า” are established for “Arabica”. However, most people are not aware of transliteration. Many variants of transliteration from English to Thai is inevitable. Fuzzy search would help in addition to adding the variants found in the EF database. Parts of the search results are depicted in Fig. 4.

50 results found for “กาแฟ Arabica”
50 Activities, 50 Emission factors

Showing items 1 - 15 of 50 total

Category	Name	Unit	Factor (kgCO ₂ e/Unit)	Reference	Last Updated	Score
กลุ่มผลิตภัณฑ์ทางการเกษตรและอาหาร	กาแฟอาราบิก้าคั่ว	kg	10.2087	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Mar 2021	139.2078
กลุ่มผลิตภัณฑ์ทางการเกษตรและอาหาร	กาแฟอาราบิก้า	kg	11.0291	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Mar 2021	46.943436
กลุ่มผลิตภัณฑ์ทางการเกษตรและอาหาร	กาแฟอาราบิก้า (ฮันท์)	kg	0.9297	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Mar 2021	46.299072
อาหาร	กาแฟคั่วอาราบิก้า 100% ดาร์ก โรสต์ 250 กรัม	200 กรัม	4 kg	มูลนิธิพลังงานไทยธรรมราชบัณฑิต, กรุงเทพฯ	Sep 2015	36.766052
อาหาร	กาแฟคั่วอาราบิก้า 100% คั่วอ่อน (ประดิมผล)	1 กก	3.05 kgCO ₂ e	บริษัท อิมเนอส์ คอร์ปอเรชั่น จำกัด, พะเยา	Feb 2023	36.64949
อาหาร	กาแฟคั่วอาราบิก้า 100% คั่วกลาง (ประดิมผล)	1 กก	3.12 kgCO ₂ e	บริษัท อิมเนอส์ คอร์ปอเรชั่น จำกัด, พะเยา	Feb 2023	36.64197
อาหาร	กาแฟอาราบิก้า 100% Medium Roast ข้างหนัก 250 กรัม	250 กรัม	2.37 kgCO ₂ e	อาซาฮีเนอส์ คอร์ปอเรชั่น จำกัด, พะเยา	Jan 2020	36.593365
อาหาร	กาแฟอาราบิก้า 100% M.Dark Roast ข้างหนัก 250 กรัม	250 กรัม	2.46 kgCO ₂ e	อาซาฮีเนอส์ คอร์ปอเรชั่น จำกัด, พะเยา	Jan 2020	36.574482
อาหาร	กาแฟคั่วอาราบิก้า 100% คั่วเข้ม (ประดิมผล) ขนาด 250 กรัม	1 กก	3.25 kgCO ₂ e	บริษัท อิมเนอส์ คอร์ปอเรชั่น จำกัด, พะเยา	Feb 2023	33.22001

Fig. 4. Search results of “กาแฟ Arabica”.

C. Case Study 3: “ถั่วเหลือง”

Fig. 5 illustrates parts of the retrieval of querying “ถั่วเหลือง” or Soybean in English. It is found in CFP database categorized in Animal feed (อาหารสัตว์), Agricultural products and food (ผลิตภัณฑ์ทางการเกษตรและอาหาร); also found in the category of Food (อาหาร) contained in Carbon Label Project dataset.

50 results found for “ถั่วเหลือง”
50 Activities, 50 Emission factors

Showing items 1 - 15 of 50 total

Category	Name	Unit	Factor (kgCO ₂ e/Unit)	Reference	Last Updated	Score
กลุ่มอาหารสัตว์	ถั่วเหลือง	kg	0.728	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019	15.051188
กลุ่มอาหารสัตว์	กากถั่วเหลือง	kg	1.024	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019	14.3124
กลุ่มอาหารสัตว์	ถั่วเหลืองฮันท์	kg	0.0969	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Mar 2021	14.261543
กลุ่มผลิตภัณฑ์ทางการเกษตรและอาหาร	ถั่วเหลืองคั่ว	kg	0.5724	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Mar 2021	13.65279
อาหาร	ถั่วเหลืองคั่วและเมล็ดถั่วเหลือง	400 กรัม	784 g	บริษัท เอ็มโอบีเอส จำกัด (มหาชน), เชียงใหม่	Sep 2014	11.948235

Fig. 5. Search results of “ถั่วเหลือง”.

D. Case Study 4: “E85”

“E85” is the compound of Gasoline and Ethanol with the ratio of 15%:85%. Observing that none of exact match when querying “E85”. The excerpt of search results is depicted in Fig. 6, matching with the synonyms, namely Gasoline (แก๊สโซลีน, ก๊าซโซลีน), น้ำมันเบนซิน (Benzene).

50 results found for “E85”
50 Activities, 50 Emission factors

Showing items 1 - 15 of 50 total

Category	Name	Unit	Factor (kgCO ₂ e/Unit)	Reference	Last Updated	Score
กลุ่มพลังงาน: เชื้อเพลิงเหลวและเชื้อเพลิงแข็ง	Gasoline (แก๊สโซลีน)	kg	0.4024	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019	38.67561
ปิโตรเลียม	ก๊าซโซลีนธรรมชาติ	kg	578 g	บริษัท ปตท. จำกัด (มหาชน), กรุงเทพฯ	Sep 2017	23.65519
ปิโตรเลียม	ก๊าซโซลีนธรรมชาติ 1 กิโลกรัม	kg	564 gCO ₂ e	บริษัท ปตท. จำกัด (มหาชน), กรุงเทพฯ	Aug 2020	21.594498
กลุ่มเครื่องจักรกลทางการเกษตร	การใช้น้ำมันดีเซลของรถ (ดีเซล 18 แรงม้า) ขับเคลื่อนรถไถนาแบบเดิน	hr	7.563	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019	21.346104
ปิโตรเลียม	ก๊าซโซลีนธรรมชาติ (โรนอค ก๊าซธรรมชาติ)	kg	391 gCO ₂ e	บริษัท ปตท. จำกัด (มหาชน), กรุงเทพฯ	Jul 2023	19.171762

Fig. 6. Search results of “E85”.

E. Case Study 5: “Sulfur”

The synonyms of “Sulfur” include ซัลเฟอร์ (transliteration), and กำมะถัน (in Thai). Fig. 7 illustrates parts of the search results consisting of the retrievals from CPF dataset (categories of Chemicals and Fuel) in addition to Carbon Label Project dataset (categories of Miscellaneous and Petroleum).

50 results found for “Sulfur”
50 Activities, 50 Emission factors

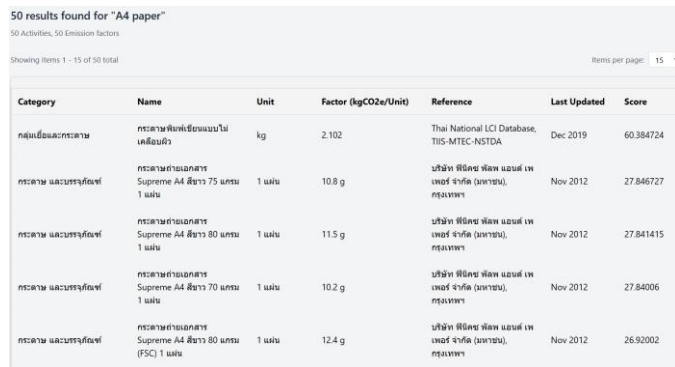
Showing items 1 - 15 of 50 total

Category	Name	Unit	Factor (kgCO ₂ e/Unit)	Reference	Last Updated	Score
อื่นๆ	ซัลเฟอร์ 99.9% บริสุทธิ์ ขนาด 5	kg	13.4 kgCO ₂ e	บริษัท ซี-เนเธอร์แลนด์ จำกัด (สำนักงานใหญ่), สระบุรี	Nov 2024	19.147892
กลุ่มผลิตภัณฑ์ (Chemicals)	Sulfur	kg	0.4529	Thai National LCI Database/MTEC	Sep 2024	17.637012
กลุ่มพลังงาน: เชื้อเพลิงเหลวและเชื้อเพลิงแข็ง	ซัลเฟอร์	kg	0.239	Thai National LCI Database, TIS-MTEC-NSTDA (with TGO electricity 2016-2018)	Dec 2019	16.379486
ปิโตรเลียม	กำมะถันขนาด 1 กิโลกรัม	kg	184 gCO ₂ e	บริษัท บางจาก คอร์ปอเรชั่น จำกัด (มหาชน), กรุงเทพฯ	Jan 2019	14.944131

Fig. 7. Search results of “Sulfur”.

F. Case Study 6: “A4 Paper”

In Thailand, “A4 paper” is common term for purchasing multi-purpose paper, or copier paper (กระดาษถ่ายเอกสาร). The synonym uncoated paper (กระดาษไม่เคลือบผิว) exists in the ecoinvent database. Fig. 8 depicts parts of matched entries, consisting of one categorized in Pulp and Paper (เยื่อและกระดาษ) found in CPF dataset, and some categorized in Paper and Packaging contained in Carbon Label Project dataset.



Category	Name	Unit	Factor (kgCO2e/Unit)	Reference	Last Updated	Score
เยื่อและกระดาษ	กระดาษพิมพ์เขียนแบบไม่เคลือบผิว	kg	2.102	Thai National LCI Database, TIIS-MTEC-NSTDA	Dec 2019	60.384724
กระดาษ และบรรจุภัณฑ์	กระดาษถ่ายเอกสาร Supreme A4 สีขาว 75 แกรม 1 แผ่น	1 แผ่น	10.8 g	บริษัท สปิธอส โพลี แอนด์ เทคโนโลยี จำกัด (มหาชน), กรุงเทพฯ	Nov 2012	27.846727
กระดาษ และบรรจุภัณฑ์	กระดาษถ่ายเอกสาร Supreme A4 สีขาว 80 แกรม 1 แผ่น	1 แผ่น	11.5 g	บริษัท สปิธอส โพลี แอนด์ เทคโนโลยี จำกัด (มหาชน), กรุงเทพฯ	Nov 2012	27.841415
กระดาษ และบรรจุภัณฑ์	กระดาษถ่ายเอกสาร Supreme A4 สีขาว 70 แกรม 1 แผ่น	1 แผ่น	10.2 g	บริษัท สปิธอส โพลี แอนด์ เทคโนโลยี จำกัด (มหาชน), กรุงเทพฯ	Nov 2012	27.84006
กระดาษ และบรรจุภัณฑ์	กระดาษถ่ายเอกสาร Supreme A4 สีขาว 80 แกรม (FSC) 1 แผ่น	1 แผ่น	12.4 g	บริษัท สปิธอส โพลี แอนด์ เทคโนโลยี จำกัด (มหาชน), กรุงเทพฯ	Nov 2012	26.92002

Fig. 8. Search results of “A4 Paper”.

V. EVALUATION

The evaluation employed a ground truth comparison methodology to assess the system's retrieval performance across different search strategies using distinct search terms.

- keyword matching
- keyword and synonyms matching
- vector search
- vector search of keyword and synonyms

Table 2 summarizes the search results, comparing the performances among the four strategies. Ground truth denotes the total number of relevant documents that should ideally be retrieved for complete coverage.

TABLE II. Comparisons of Keyword and Hybrid search.

Query input	Search				ground truth
	Keyword	Keyword + synonyms	Hybrid	Hybrid + synonyms	
ก๊าซหุงต้ม	8	20	10	20	20
กาแฟ Arabica	19	24	19	25	25
Arabica	0	23	1	23	25
ถ้วยเหลือง	163	163	163	163	163
แก๊สโซลีน	8	13	8	14	14
Gasoline	7	13	8	13	14
A4 Paper	4	5	4	5	5
E85	0	12	0	13	14
Sulphur	0	4	1	4	4
Sulfur	1	4	1	4	4

The keyword searching with “Arabica” mismatched because none of the word “Arabica” exists in the database. While keyword matching with “กาแฟ Arabica”, the search result returns 19 records because the results matched with the word “กาแฟ”.

The search results of “ถ้วยเหลือง” are overwhelming. Probably, more specific term is suggested, and the semantic search would better suit in this scenario.

The search results of querying “แก๊สโซลีน” and “Gasoline” are quite indifferent as the result of cross-lingual search applied.

The reason why keyword searching with “Sulphur” returns null is there is none of item named “Sulphur”. The database merely contains “Sulfur”. However, “Sulphur” and “Sulfur” considered alternatives, they are added into synonyms.

It is observed that using synonyms in either keyword search or hybrid search achieved greater matches compared to none of synonyms applied. The hybrid search using synonyms outperformed overall strategies due to the fact that semantic search goes beyond keyword matching to understand the meaning and intent behind a search query, thus providing more relevant results.

The experimental results also reported the finding that the performance of keyword matching is close to that of hybrid search. In the early stage of this work, the query input simply contains words or phrase in Thai and English languages. The cross-lingual search accompanied with synonyms provides satisfactory results of keyword matching. However, the fact that the user input is likely the common words used in daily life that differs the entries contained in the emission factor databases. Hence, semantic search would be effective and useful for further direction.

REFERENCES

- [1] Thai National LCI Database, by Technology and Informatics Institute for Sustainability, 2025. Available: <https://www.nstda-tiis.or.th/lci-database/status-of-national-lci-database/>
- [2] AR5, by Intergovernmental Panel on Climate Change (IPCC), 2025. Available: <https://ipcc-data.org/>
- [3] Ecoinvent Database, by Ecoinvent, 2025 Available: <https://ecoinvent.org/>
- [4] Emission Factor (CFO, CFP), by Thailand Greenhouse Gas Management Organization (TGO), 2025. Available: <https://thaicarbonlabel.tgo.or.th/index.php?>
- [5] R. Changala, S. Barathi, R. Subhashini, J. C. Lawrance, and B. K. Bala, "Implementing Cross-Lingual Information Retrieval Systems to Enhance Resource Accessibility in English Language Learning," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2024: IEEE, pp. 1-6.
- [6] M. Aliramezani, E. Doostmohammadi, M. H. Bokaei, and H. Sameti, "Persian sentiment analysis without training data using cross-lingual word embeddings," in *2020 10th International Symposium on Telecommunications (IST)*, 2020: IEEE, pp. 78-82.
- [7] T. Grainger, D. Tumbull, and M. Irwin, *AI-Powered Search*. Simon and Schuster, 2025.
- [8] English Thesaurus, by Cambridge Dictionary, 2025. Available: <https://dictionary.cambridge.org/>
- [9] Thai dictionary, by Royal society of thailand, 2025. Available: <https://dictionary.orst.go.th/>
- [10] Emission Factor (Carbon Footprint for Organization), by Thailand Greenhouse Gas Management Organization (TGO), 2025. Available: <https://thaicarbonlabel.tgo.or.th/index.php?lang=TH&mod=YjNKblXNXBlbUYwYVc5dVgyVnRhWE56YVc5dQ>
- [11] Emission Factor (Carbon Footprint of Product), by Thailand Greenhouse Gas Management Organization (TGO), 2025. Available: <https://thaicarbonlabel.tgo.or.th/index.php?lang=TH&mod=Y0hKdIpIVmpkSE5mWlcxcGMzTnBiMjQ9>
- [12] Carbon Label Project, by Thailand Greenhouse Gas Management Organization (TGO), 2025. Available: <https://thaicarbonlabel.tgo.or.th/index.php?lang=TH&mod=WTJGMFIx>



eHZadz09&action=Y0c5emRBPT0§ion=0&industry=0&style=_TABLE&sorting=_ASC&year=0&quarter=0

- [13] M. Konda, Elasticsearch in action. Simon and Schuster, 2024.
- [14] S. S. Badhya, A. Prasad, S. Rohan, Y. Yashwanth, N. Deepamala, and G. Shobha, "Natural language to structured query language using

elasticsearch for descriptive columns," in 2019 4th international conference on computational systems and information technology for sustainable solution (CSITSS), 2019: IEEE, pp. 1-5.