Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 80 (2019) 643–648

www.elsevier.com/locate/procedia

26th CIRP Life Cycle Engineering (LCE) Conference

A collaboration platform for enabling industrial symbiosis: Towards creating a self-learning waste-to-resource database for recommending industrial symbiosis transactions using text analytics

Zhiquan Yeo[a,b*], Jonathan Sze Choong Low[b], Daren Zong Loong Tan[b], Si Ying Chung[b], Tobias Bestari Tjandra[b], Joshua Ignatius[a]

[a]*WMG, University of Warwick, Coventry, CV4 7AL, United Kingdom*

[b]*Singapore Institute of Manufacturing Technology, 2 Fusionopolis Way, Singapore 138634*

* Corresponding author. Tel.: +44 (0)7864 660 884/+65 6319 4463. *E-mail address:* z.yeo@warwick.ac.uk

Abstract

Industrial Symbiosis (IS) adopts a collaborative approach, which aims to re-channel resources – traditionally considered spent and non-productive – towards alternative value-adding pathways. Empirically, the concept of IS has been rapidly implemented in practice through a facilitated approach, whereby businesses are engaged and "match-made" via a facilitating body. While recommending alternative pathways for companies to establish IS-based transactions is a long-standing practice, recent technological advancement has shifted the nature of this task from one that is based purely on human intellect and reasoning, towards one which leverages intelligent recommendation algorithms to provide relevant suggestions. Traditionally, these recommendation engines rely on manually populated knowledge bases that are not only labor-intensive to build but also costly to maintain. This work presents the creation of a self-learning waste-to-resource database supporting an IS recommendation system by utilizing text analytics techniques. We further demonstrate its practical application to support IS facilitating bodies in their core activity.

Peer-review under responsibility of the scientific committee of the 26th CIRP Life Cycle Engineering (LCE) Conference.

Keywords: industrial symbiosis; recommender; machine learning; text analytics

1. Introduction

Industrial Symbiosis (IS) entails the notion of inter-firm cooperation, working towards the systemic goal of improved economic and ecological performance [1]. These inter-firm cooperation often translate to novel sourcing of process inputs and value-added destinations for non-productive outputs, leading to mutually beneficial transactions [2] involving materials, energy, water, by-products [3], assets, logistics, expertise and knowledge [2]. Beneath these transactions underlie the processes of shared knowledge creation and eco-innovation, enabled through a network of participating organizations [2].

In practice, establishing IS transactions requires a process of facilitation and coordination [4] and various barriers (e.g. informational, technical, economic) have been recognized in current literature [5]. One of the key barriers identified by various authors – and the focus of this discussion – is the *information barrier* [6], [7]. Specifically, it refers to the deficiency of information exchange, resulting in the lack of visibility of the market in the aspects of extant potential demand and supply of underutilized resources (i.e. waste) [7]. This in turn causes difficulty in finding value-adding pathways for waste products – an essential element in creating IS-based transactions.

Peer-review under responsibility of the scientific committee of the 26th CIRP Life Cycle Engineering (LCE) Conference.
10.1016/j.procir.2019.01.015

Addressing the information barrier, researchers and practitioners have developed various purpose-built IS tools [8]. These tools aim to enable and manage the information flow that fosters the formation of IS-based linkages between companies. Specifically, the common feature of these tools is the function to provide decision makers recommendations on suitable matches of waste streams to their relevant receiving processes, technologies and/or companies [8]–[10]. However, the lack of usability resulted in most of these tools to be fallen from use [8].

A critical pitfall for such data intensive IS tools is the foremost requirement to populate the tools with required data; without which the tools (and their functions) only act as a framework without practical use [9]. This is akin to the "cold-start problem" in information systems [11], referring to the situation whereby little or no inferences can be drawn if only sparse or no data is available for analysis at the beginning. This diminishes the usefulness of the system as it is unable to achieve its intended goal without sufficient data. Therefore, whilst IS tools and their embedded recommendation algorithms were designed in the past, most if not all of these tools were designed with the assumption that data will naturally be populated over time and they will gain the momentum to sustain their use thereafter. The low number of tools empirically observed in use today [8], however, suggests that this key assumption needs to be revisited and further addressed moving forward.

In the recent three years, there has been increased development activities and sophistication of the next generation of IS tools. The recent trends in IS tool development include cloud-based platforms featuring dynamic databases and recommendation algorithms that promote collaboration by matching traditional and non-traditional industrial waste streams with novel product and revenue opportunities [12]–[19]. As these newer tools are still in their development phase, their efficacy has yet to be observed. However, in order to avoid similar fate as their predecessors, the authors posit the issue of data (or the lack of it) deserves equal attention as its accompanying recommendation algorithms when developing these next-generation IS decision support tools.

2. System Architecture of Collaboration Platform

Throughout our ongoing effort towards an IS collaboration platform [17], [19], we have made use of a system architecture which allows us to integrate our subsystems that weaves through the fundamental layers: data, logic and the user interface (UI) shown in Figure 1.

The vertical groupings represent the different subsystems that the collaboration platform encompasses to allow for end-to-end resource matches. It comprises various integrated solutions that starts off by enabling firms to participate with the relevant detail inputs. The resource inputs will be used to query the graph database which returns the feasible options for conversion or substitution of resource. Through these possibilities, the industrial symbiosis matchmaking subsystem will attempt to matchmake firms based on a set of algorithms. With this integrated pipeline of solutions, firms will be able to evaluate these results through the simulation to provide another level of quantification to determine the actual feasibility of these exchanges through features such as economic and supply-demand fulfilment viability.

From the firm users' perspective, the UI will provide the layer of presentation to ease the interaction with the platform. The logic layer comprises the computations required to process information, the algorithms and acts as the gateway to external knowledge sources to support our data-driven analytic approaches. The data layer contains information of participating firms, resource details and knowledge on feasible resource matches.

Alongside today's agile developments and evolving requirements, we are continuing to adapt and this system architecture provides the fundamental building blocks for us to scale. In the sections that follow, this paper will focus on the process of enabling the self-learning waste-to-resource graph database (W2RDB).

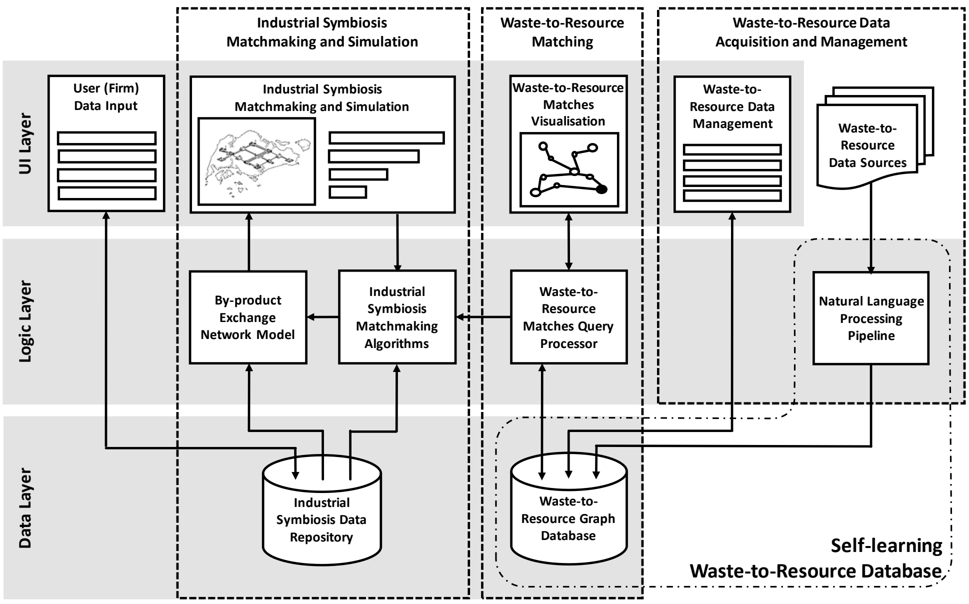

Figure 1: UI layer, Logic layer and Data layer of the collaboration platform

3. Self-learning waste-to-resource graph database (W2RDB)

The concept of knowledge databases (or repositories) is not new to the field of IS, whereby both publicly available [22], [23] and proprietary databases [24] exist. These knowledge databases serve to provide the background knowledge required to suggest possible IS matches given a set of available streams and processes. The use of knowledge database assumes that knowledge required for supporting IS can be found within the sources such as academic literature, case studies/reports and patents. Therefore, knowledge describing IS stream and process relations are typically extracted manually and represented as tabular formats (e.g. [22], [23]). However, in practice, the publicly available databases lack continuous efforts in updating the knowledge libraries, failing to continuously capture new knowledge, while the proprietary counterparts lack transparency and serves the benefit of their exclusive owners. Therefore, in our research, we aim to improve the current state-of-the-art by contributing towards a method for creating a self-learning W2RDB that minimizes costly time and effort that inhibits the long-term maintenance of such knowledge repositories.

3.1. Constructing the W2RDB conventionally

In order to arrive at our goal of creating a self-learning W2RDB, we first established a deeper understanding on the conventional process of constructing such a knowledge database, previously described in [19].

For the purpose of proof-of-concept, a collection of scientific papers embedding knowledge about the derivation of useful resources from food waste materials were collected by searching scientific journal databases with relevant keywords such as “food waste” and “food waste valorization”. Papers cited by the review paper written by Ravindran and Mirabella [21], [22] were also added into the collection. The gleaned information was then manually extracted from the papers and filled into the structure as depicted in Figure 2. The data structure was further transformed into a form suitable for the interoperability with the IS collaboration platform.

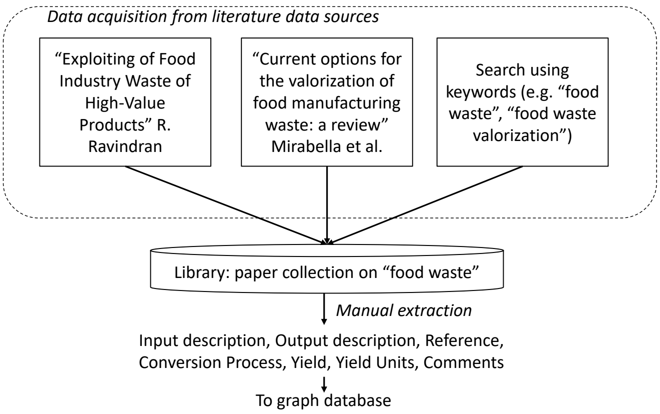

Figure 2: Manual database construction process

The collaboration platform embraces polyglot persistence by using multiple data storage technologies to handle the varying data storage requirements. This allows us to provide good data storage fit in the form of a combination of both relational databases for knowledge sources, inputs and a graph database specifically for the W2RDB.

The graph database has been selected as the data model for our W2RDB as it is most representative for our use case with highly interconnected data and suits the way we query resource matches compared to the typical relational databases [19]. The graph database is modelled with three distinct node types; resource, process and process category. The process node acts akin to how a cooking recipe would function. It captures exclusive information and is unique to every resource match. By decoupling the process categories, we are able to differentiate resource matches while still having a relation to the process categories that helps manage data redundancy. Each node can also have any number of properties which makes them more flexible to cater to the diversity of resource information. With this data model we are able to provide more complex indirect and process-based waste-to-resource queries while providing scalability and flexibility for any unstructured and diverse data.

3.2. Challenges with scaling up the W2RDB

The prototype of the W2RDB previously discussed by Low *et al.* [19] depicted its operation in various use cases and demonstrated how IS recommendations can be generated based on the knowledge available. In this paper, we focus on how the data acquisition of this W2RDB can be scaled up economically.

Due to the diverse nature of waste, an equally diverse set of knowledge is required to be included in the database. Consider the case if only streams and technologies related to food waste is captured by the W2RDB, when the system encounters a query related to a different category of stream such as plastic waste, a null result will be returned as the knowledge related to plastic waste does not exists within the W2RDB.

The practical challenge of scaling the W2RDB is the ever-growing set of data required to be sieved through and transformed for use. According to Bornmann [20] and Mortenson and Vidgen [21], the growth rate of scientific output is 8-9% per annum, which translates to doubling every nine years. Davis and Aid [22] suggested that the rate of converting raw data into machine sensible information is limited by the human’s ability during the conversion process – termed as “encoding bottleneck”. Therefore, the rapid growth of the amount of knowledge available for encoding, coupled with the human limits makes scaling up and maintaining the W2RDB a costly endeavor by conventional means. In the prototype database, the authors took approximately six months to gather 742 data entries.

To overcome the “encoding bottleneck”, Davis and Aid [22] applied a word vector based approach that transfers the burden of encoding raw data from the human to machine. In their work, they exhibited that word vector is a promising step for encoding knowledge as it endows machines with the ability to represent semantics that is machine process-able and produces query outputs that are humanly sensible. In essence, word vectors represent language lexicons in vector space and it is empirically observed to represent semantic/syntactic relatedness and relations [23]. For instance, the examples provided by Mikolov *et al.* [23] exhibited the machine's ability to represent human verifiable relations in the general knowledge domain such as capital-cities, currency-country, among others.

Nevertheless, applying word vectors to scale up the W2RDB, while essential, is insufficient. In practice, training a word vector model produces a lookup table of vectors for lexicons derived from a neural network based training process. In other words, this step creates machine representable features for individual lexicons derived from the raw knowledge sources. However, in order to make (human) sense out of the machine represented features, significant manual effort is still required to sieve through the huge vocabulary collection of lexicons to query the model. Put into the practical perspective, requiring humans to manually search through a vocabulary size in the order of 10^5-10^6 (depending on the size of training corpus) requires substantial amount of effort and rapidly becomes impracticable when dealing with even larger corpus (e.g. Google News Corpus [24] comprises 3 million words and phrases). This poses a major challenge in applying the concept of word vectors to establish an autonomous system to extract required knowledge from texts as word vectors alone does not eliminate the need for substantial human effort to make sense of the output.

3.3. Natural Language Processing pipeline for knowledge extraction

To overcome the mentioned challenge in enabling a self-learning W2RDB, we adapted and extended the work of Davis and Aid [22] and constructed a Natural Language Processing (NLP) pipeline for extracting the desired knowledge, beginning from data acquisition to producing outputs suitable for interfacing with the graph database infrastructure described by Low *et al.* [19]. Figure 3 illustrates the components constituting the NLP pipeline.

The first component is responsible for acquiring data from relevant web-based knowledge sources such as the academic literature, patent databases and grey literature sources (e.g. IS case reports). This component connects to various data plugs and aggregates data into a central location for processing.

The second component serves to interpret various data formats (e.g. PDF, XML) to extract only relevant portions of the original data document. For instance, only the main body of the documents is extracted while ancillary document components such as page numbers and document formatting information are discarded.

In the text pre-processing stage, noun phrases are joined and represented as a single lexicon so that it enables the machine to learn the noun phrases as a whole rather than at the individual word level. For example, "*waste cooking oil*" is joined to form "*waste_cooking_oil*". In the case of the former, "*waste*", "*cooking*" and "*oil*" will be treated as independent lexicons and will be learnt by the machine as three independent word meanings. In the case of the latter however, it enables the machine to recognize the phrase as a whole and it will learn the word meaning as a single entity. This is necessary as the constituent words of phrases convey different semantics from their respective phrases in reality. Noun phrases was selected taking into consideration that waste streams, processes and technologies are frequently represented by noun or noun phrases in text by observation.

The fourth step involves generating word vectors that create a machine representation of lexicons in the vocabulary derived from all the knowledge documents. As discussed in section 3.2, direct human interpretation of the generated word vectors is not possible. Subsequent steps however will utilize results of word vectors to derive (human) sensible information in an autonomous way.

The named entity recognition (NER) step comprises a customized module that examines documents to identify entities comprising the property of "*process*" and "*resource*". We define "*process*" as any technology or process that transforms resource input into valuable resource output, and "*resource*" as any stream that can be accepted by any "*process*". As terms with related meaning are spatially correlated, empirically observed by Mikolov *et al.* [25], a classification approach was used to segregate *process* and *resource* terms from the detected noun chunks.

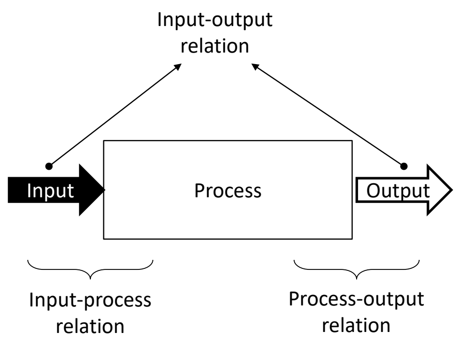

Figure 4: Relations extraction to link named entities to their relative position with respect to a particular process

Finally, the relation extraction step of the pipeline examines documents to extract three types of relations – namely "*input-process*", "*process-output*" and "*input-output*" (Figure 4) – to construct a possible resource valorization pathway. The relation extraction step examines the entity pairs as well as surrounding word context and determines the relation confidence level. These three relations are selected primarily for a triangulation process to confidently determine the existence of a process and its corresponding inputs and outputs. For instance, if the following relations are detected: (a) input I

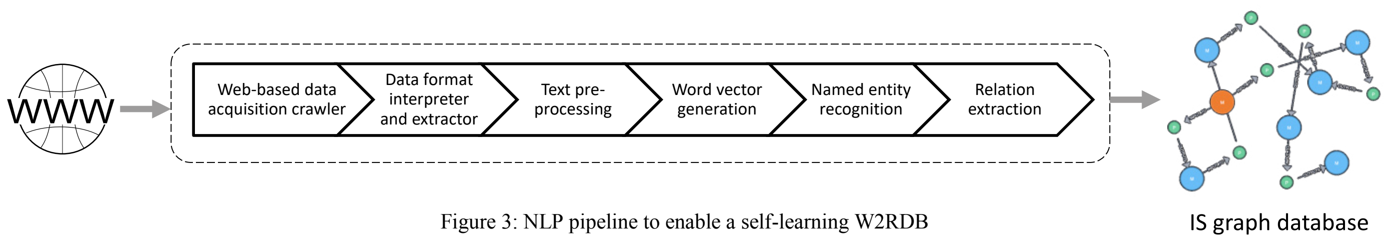

Figure 3: NLP pipeline to enable a self-learning W2RDB

and process P are related by "*input-process*" relation, (b) process P and output O are related by "*process-output*" relation, (c) input I and output O are related by "*input-output*", we can deduce that I, P and O exist as a continuous sequential pathway. The final outputs of the NLP pipeline are possible stream conversion pathways represent by the tuple structure of (*input, process, output*).

4. Demonstration

This section demonstrates the interim outputs produce at each step of the prototype of the NLP pipeline implemented in Python 3.

The outputs of the web-based data acquisition crawler are documents represented in various formats that exist "in the wild" which requires much processing and refining for suitable use. Figure 5 is an example of a retrieved article [26] from Elsevier API [27] to illustrate the general nature of the data retrieved.


```
from Algae </dc:title><prism:publicationName>Procedia Engineering
</prism:publicationName><prism:aggregationType>Journal</prism:aggregat
</prism:issn><prism:volume>42</prism:volume><prism:startingPage>231</p
</prism:endingPage><prism:pageRange>231-238</prism:pageRange><dc:forma
2012-12-31</prism:coverDate><prism:coverDisplayDate>2012</prism:coverD
© 2012 Published by Elsevier Ltd.</prism:copyright><prism:publisher>Pu
</prism:publisher><prism:issueName>CHISA 2012</prism:issueName><dc:cre
</dc:creator><dc:creator>Cvengrošová, Z.</dc:creator><dc:creator>Rimar
Buzetzki, E.</dc:creator><dc:creator>Mikulec, J.</dc:creator><dc:creat
</dc:creator><dc:description>
                    Abstract

                         Fatty acid methyl esters (FAME) are used as alternat
                         renewable sources. The attention is focused on the m
```


Figure 5: Sample of crawled raw data in XML format (with XML tags and body content)

The acquired raw data are relayed to the data format interpreter, whereby the appropriate data conversions are applied to extract only relevant parts of the raw data. In our case, our interest is to analyze the main content of the documents (an example shown in Figure 6).

```
water, inorganic  salts (N, P, K) and temperature in the
fix CO 2  from three different  sources: atmosphere, dis
from soluble carbonates [2]. Microalgal  biomass contain
weight [3]. Producing 100 t of algal biomass fixes  roug
species, microalgae produce many different kinds of lipi
oils [4  6]. In general, many algae species have the oil
```

Figure 6: The extracted plaintext sample of the XML document

The text of key interest subsequently undergoes pre-processing to identify and join noun phrase into a single chunk. Figure 7 shows the sample output of this step of the example document.

```
effectiveness of energy_conversion to biomass due to
metabolic_functions.  Moreover, they fix waste_CO 2
destroy the algae_cell in order to  obtain the oil.
FAME_preparation from received_algae_oils are presen
Nannochloropsis_and_Chlorella_microalgae provide val
biofuel_production. Characteristics of prepared_FAME
```

Figure 7: Text pre-processing output joining noun chunks

Following that, the pre-processed text is streamed to word vector generation, whereby a neural network will be learnt to derive the word vectors for the lexicons found in the text stream. At the end of this step, the output is fundamentally a dictionary of lexicons with their associated multi-dimensional vectors.

The NER will further process the text to detect named entities and assign them the appropriate labels. For instance, "*Nannochloropsis_and_Chlorella_microalgae*" will be assigned "*resource*" while "*biofuel_production*" will be assigned "*process*".

Finally, the relation extraction will analyze the text to identify the appropriate relationships being conveyed. To illustrate the concept, consider the following text "Using *microalgae* in the *biofuel_production* will not compete with the production of food, fodder and other products from crops". The term "*microalgae*" was identified as a "*resource*" while "*biofuel_production*" was identified as a "*process*" in the previous step. The relationship extractor will then analyze the surrounding text (i.e. context) around those identified as "*resource*" and "*process*". In this case, the relation between "*microalgae*" and "*biofuel_production*" is of the type "*input-process*". Similarly, the text "producing *biofuels* from *microalgae*" will lead to the identification that "*microalgae*" and "*biofuel*" are related by the "*input-output*" relation. Additionally, the text "the *biofuel* produced by the *biofuel_production_process* has higher yield" will classify the relation between "*biofuel*" and "*biofuel_production_process*" as "*process-output*" relation. A reconciliation process will be performed to link the three relations. For instance, in Table 1, based on the relations and entities identified, the final output tuple (*microalgae*, *biofuel_production/biofuel_production_process*, *biofuel*) will be produced, which conveys that microalgae can be accepted by the biofuel production process to produce biofuel as the output. This tuple conveys the knowledge required by the graph database and hence will be piped to it. Subsequently, this graph database will be integrated with the matchmaking algorithms discussed by Raabe *et al.* [17], whereby firms will be matched according to their appropriate resource specifications. In our preliminary prototype, the NLP pipeline took approximately five days to run on a system with 2.60 GHz quad core CPU and 16 GB of RAM.

Table 1: Entities and relations extracted

Entity 1	Entity 2	Relation type
microalgae	biofuel_production	input-process
biofuel_production_process	biofuel	process-output
microalgae	biofuel	input-output

5. Conclusion

In this paper, we focus on establishing a NLP pipeline capable of processing large quantities of text in order to extract information to support the population of an IS knowledge repository (named as W2RDB). The W2RDB is subsequently utilized as a background knowledge reference to recommend viable alternative conversion pathways for waste streams for the IS collaboration platform. This research contributes towards a cost-effective method for constructing an instance of a knowledge repository required by the field of IS to conduct matchmaking among firms – a core characteristic activity of facilitating and fostering IS networks around the world.

While this current work is still in progress, we have successfully demonstrated a proof-of-concept. Further refinement to the individual modules of the NLP pipeline is ongoing at the time of writing and we expect improved performance over time.

The critical determinants of the quality of output tuples are: (i) quality of knowledge sources, (ii) quality of pre-processing, (iii) word vector training parameters, (iv) NER and classification performance, and (v) relation extraction technique. Our future work includes focusing on addressing the mentioned determinants of the quality of output tuples and to perform validation of the results to obtain estimates of performance. Eventually, the proposed technique will be incorporated into the IS collaboration platform, enabling its capability to be self-learning, enriched by the perpetually growing knowledge sources.

6. Acknowledgements

We would like to thank Michael Mortenson (University of Warwick), Chris Davis (University of Groningen) and Graham Aid (Ragn-Sells AB) for their valuable inputs on NLP and word vectors. We would also like to thank Keane Looi Yi Xian, Yim Chia Hui and Kenrick Goh Chung Tat (National University of Singapore) for their contributions towards parts of the technical implementation.

References

[1] M. Chertow and J. Park, "Scholarship and Practice in Industrial Symbiosis: 1989–2014," in *Taking Stock of Industrial Ecology*, Springer, 2016, pp. 87–116.

[2] D. R. Lombardi and P. Laybourn, "Redefining Industrial Symbiosis: Crossing Academic-Practitioner Boundaries," *J. Ind. Ecol.*, vol. 16, no. 1, pp. 28–37, Feb. 2012.

[3] M. Chertow, "Industrial symbiosis: Literature and taxonomy," *Annu. Rev. ENERGY Environ.*, vol. 25, pp. 313–337, 2000.

[4] M. Chertow and J. Ehrenfeld, "Organizing Self-Organizing Systems: Toward a Theory of Industrial Symbiosis," *J. Ind. Ecol.*, vol. 16, no. 1, pp. 13–27, Feb. 2012.

[5] A. Golev, G. D. Corder, and D. P. Giurco, "Barriers to Industrial Symbiosis: Insights from the Use of a Maturity Grid," *J. Ind. Ecol.*, vol. 19, no. 1, pp. 141–153, 2015.

[6] M. Chertow, "'Uncovering' industrial symbiosis," *J. Ind. Ecol.*, vol. 11, no. 1, pp. 11–30, 2007.

[7] D. Gibbs, "Trust and Networking in Inter-firm relations: the Case of Eco-Industrial Development," *Local Econ.*, vol. 18, no. 3, pp. 222–236, 2003.

[8] G. B. Grant, T. P. Seager, G. Massard, and L. Nies, "Information and communication technology for industrial symbiosis," *J. Ind. Ecol.*, vol. 14, no. 5, pp. 740–753, 2010.

[9] J. Brown, D. Gross, and L. Wiggs, "The MatchMaker! System: Creating Virtual Eco-Industrial Parks," *Yale F&ES Bull.*, no. 106, pp. 103–136, 1997.

[10] P. Laybourn and M. Morrissey, "The Pathway To A Low Carbon Sustainable Economy," *Natl. Ind. Symbiosis Program.*, vol. 44, no. 0, pp. 2009–2010, 2009.

[11] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," *Proc. 25th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '02*, no. August, p. 253, 2002.

[12] SHAREBOX, "SHAREBOX – SECURE SHARING," 2017. [Online]. Available: http://sharebox-project.eu/. [Accessed: 11-Sep-2017].

[13] MAESTRI, "Home - MAESTRI - Energy and resource management systems for improved efficiency in the process industries," *MAESTRI Project Homepage*, 2017. [Online]. Available: https://maestri-spire.eu/. [Accessed: 27-Oct-2017].

[14] LifeM3P, "Life M3P – Material Match Making Platform," *LifeM3P Project Homepage*, 2017. [Online]. Available: http://www.lifem3p.eu/en/. [Accessed: 27-Oct-2017].

[15] The Materials Marketplace, "The Materials Marketplace," *The Materials Marketplace Website, US BCSD*, 2016. [Online]. Available: http://materialsmarketplace.org/. [Accessed: 27-Oct-2017].

[16] B. Song *et al.*, "A big data analytics approach to develop industrial symbioses in large cities," in *Procedia CIRP*, 2015, vol. 29, pp. 450–455.

[17] B. Raabe *et al.*, "Collaboration Platform for Enabling Industrial Symbiosis: Application of the By-product Exchange Network Model," in *Procedia CIRP*, 2017, vol. 61, pp. 263–268.

[18] B. Song, Z. Yeo, P. Kohls, and C. Herrmann, "Industrial Symbiosis: Exploring Big-data Approach for Waste Stream Discovery," in *Procedia CIRP*, 2017, vol. 61, pp. 353–358.

[19] J. S. C. Low *et al.*, "A Collaboration Platform for Enabling Industrial Symbiosis: Application of the Database Engine for Waste-to-Resource Matching," in *Procedia CIRP*, 2018, vol. 69, no. May, pp. 849–854.

[20] L. Bornmann, "Growth Rates of Modern Science : A Bibliometric Analysis Based on the Number of Publications and," pp. 1–28, 2015.

[21] M. J. Mortenson and R. Vidgen, "A computational literature review of the technology acceptance model," *Int. J. Inf. Manage.*, vol. 36, no. 6, pp. 1248–1259, 2016.

[22] C. Davis and G. Aid, "Machine Learning Assisted Industrial Symbiosis: Using Neural Networks and Big Data to Identify Valuable Material Substitutions," in *ISIE-ISSST 2017: Science in Support of Sustainable and Resilient Communities*, 2017.

[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," pp. 1–12, 2013.

[24] Google Code Archive, "Google Code Archive - Long-term storage for Google Code Project Hosting.," *Google Code Archive*, 2013. [Online]. Available: https://code.google.com/archive/p/word2vec/. [Accessed: 11-Sep-2018].

[25] T. Mikolov, H. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," in *Advances in neural information processing systems*, 2013, pp. 319–333.

[26] A. Pandey, D. J. Lee, Y. Chisti, and C. R. Soccol, "Biofuels from Algae," *Biofuels from Algae*, vol. 42, no. August, pp. 1–338, 2013.

[27] Elsevier, "Elsevier Developer Portal," 2018. [Online]. Available: https://dev.elsevier.com/. [Accessed: 12-Sep-2018].