UNIVERSITÀ DI PISA

DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA DEI SISTEMI DEL TERRITORIO E DELLE COSTRUZIONI

RELAZIONE PER IL CONSEGUIMENTO DELLA LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE

The future of Circular Economy: extracting and mapping skills associated with green professional profiles.

SINTESI

RELATORI

Prof. Ing. Gualtiero Fantoni, Ph.D.
Dipartimento di Ingegneria Civile e Industriale

Dr. Filippo Chiarello, Ph.D.
Dipartimento di Ingegneria dell'Energia dei Sistemi del Territorio e delle Costruzioni

Dr. Vito Giordano
Dipartimento di Ingegneria dell'Informazione

IL CANDIDATO

Francesca Costa
f.costa2@studenti.unipi.it

Sessione di Laurea Magistrale del 17/02/2021

The future of Circular Economy: extracting and mapping skills associated with green professional profiles.

Francesca Costa

Sommario

La Circular Economy (CE) è ormai un tema molto trattato sia a livello accademico che nel dibattito pubblico. Il paradigma della CE mira ad ottenere sistemi orientati a non utilizzare le risorse in esaurimento della terra e a garantire che i flussi di materiale ed energia rientrino continuamente in essi al fine di ridurre al minimo gli sprechi.

Ci sono diversi ostacoli nell'applicazione pratica dei concetti della CE, tra i quali la mancanza di una classificazione chiara e ampiamente accettata dei temi legati ad essa e di quali competenze e profili professionali impatterà. Infatti, sebbene esistano diversi database di profili professionali, non vi è alcun riconoscimento a livello europeo di quali siano da considerare green.

In questa tesi abbiamo affrontato queste problematiche utilizzando strumenti di analisi automatica del testo e abbiamo raggiunto due risultati: (i) partendo da pubblicazioni scientifiche, abbiamo ottenuto una suddivisione della CE in sotto temi; (ii) incrociando i temi ed il database europeo delle competenze e delle professioni (ESCO), abbiamo estratto quelli che possono essere considerati profili professionali legati alla CE, fornendo uno strumento utile alle aziende che vorranno adattarsi al nuovo paradigma.

Abstract

The Circular Economy (CE) is by now a highly treated topic both academically and in public debates. CE paradigm aims to safeguard the Earth resources and to ensure that the flows of material and energy continuously re-enter the system. In this way, CE minimises waste.

There are several obstacles in the practical application of CE: the most important ones are the lack of a standard classification of CE topics and the uncertainly of how the new paradigm will impact skills and job profiles. Although, there are several job and skills taxonomies, there is not a European agreement of which profiles are "green".

In this thesis, I faced these problems using automatic text analysis tools, achieving two outcomes: firstly, starting from scientific publications, I obtained a subdivision of CE into sub-topics; secondly, by crossing the topics and the European database of skills and profiles (ESCO), I extracted what can be considered CE job profiles, giving a tool to companies wanting to adapt to the new paradigm.

1. Introduction

A Circular Economy is a novel approach to economic development designed to focus on the environment sustainability in order to benefit business and society. In contrast to the 'take-make-waste' linear model, Circular Economy is regenerative, gradually decoupling growth from the consumption of finite resources[1]. Circular Economy is an increasingly important topic in the last year, in fact the Future of Jobs Report 2020 of the World Economic Forum states that the "set of emerging professions reflects the adoption of new technologies and increasing demand for new products and services, which are driving greater demand for green economy jobs"[2]. Given the broadness of CE topics, I can find at least three levels of abstraction in the application of it:

1. the macro level includes activities developed at a city, province, region or national level. It aims to promote a recycling oriented society;
2. the meso level describes an inter-firm level within geographic proximity. It includes Industrial Symbiosis and Eco-industrial parks;
3. the micro level focuses on single firm activities or on consumers[3].

In addition, the approaches to the new paradigm differ from one continent to another. For example, the transition towards CE in Europe is following a bottom-up approach (from the initiatives of environmental organizations, civil society and NGOs)[4]. In China CE is a direct outcome of the national political strategy.[4]

For these reasons several problems arise:

- lack of a univocal paradigm that includes CE concepts and prerogatives[3];
- lack of a standard subdivision of CE areas[5];
- ignorance of the knowledge and training plans necessary in future for CE professional figures[6].

[1] https://www.ellenmacarthurfoundation.org/circular-economy/what-is-the-circular-economy

[2] https://www.weforum.org/reports/the-future-of-jobs-report-2020

[3] Merli, R., Preziosi, M., Acampora, A. (2018). How do scholars approach the circular economy? a systematic literature review. Journal of Cleaner Production, 178: 703-722.

[4] Ghisellini, P., Cialani, C., Ulgiati, S. (2016). A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. Journal of Cleaner Production, 114: 11- 32

[5] Cottafava, D., Ascione, G.S., Allori, I., Chiarello, F., (2019). Circular Economy: new paradigm or just relabelling? A quantitative text and social network analysis on wikipedia webpages.

[6] Burger, M., Stavropoulos, S., Ramkumar, S., Dufourmont, J., van Oort, F. (2019). The heterogeneous skill-base of circular economy employment. Research Policy. 48 (1), pp. 248-261

I started from CE state of art. My aim was to understand how CE research fields evolved in the last decade; I did so analysing CE scientific publications using Text Mining techniques. I chose scientific papers that enable to anticipate the needs of industry and society[7]. In order to obtain the training plans (third issue mentioned above), it is necessary to understand which job profiles and skills are CE related. To sort out that, I started from already existing skills and profiles taxonomies.

To summarise, the main goals of this thesis were: (i) creating a Circular Economy subdivision into topics; (ii) creating a mapping of CE professional figures and green skills.

The results I achieved were: the subdivision of Circular Economy macro-topic into 7 sub-topics and the extraction of 85 CE professional profiles.

2. State of art

The literature review provided in this section is focused on the history of Circular Economy, the existing topic classifications and the analyses already carried out on green skills.

The economic model that our society has always followed is called the "linear model" or the "take, make, dispose model"[8]. This approach is based on the extraction of raw materials and the mass use of resources. It was implemented since the early days of modern society, as resources seemed limitless and there was no need to worry about safeguarding the planet[9]. Nowadays, the linear model has led to such a planet exploitation as to threaten natural ecosystems essential for the survival of man on Earth[4]. For these reasons, the world economy should tend towards a new model, namely Circular Economy. CE is an economic system capable of regenerating itself through reuse of all resources (raw materials, energy or space)[8]. The growing number of CE publications, as showed in *Figure 1*, is an evidence of the increasing attention paid to the phenomenon.

[7] Chiarello, F., Belingheri, P., Bonaccorsi, A., Martini, A., Fantoni, G., (2021), "Value Creation in Emerging Technologies through Text Mining: The Case of Blockchain", Technology Analysis & Strategic Management

[8] Lieder, M., Rashid, A. (2016). Towards circular economy implementation: a comprehensive review in context of manufacturing industry. Journal of cleaner production, 115, pp. 36-51

[9] Lacy, P., Rutqvist, J., Lamonica, B. (2015). Waste to Wealth: The Circular Economy Advantage. Palgrave Macmillan.

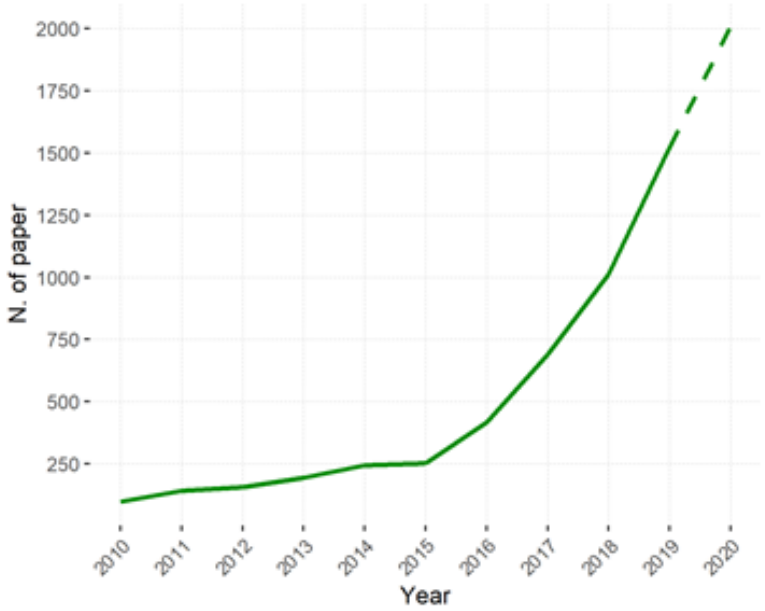

Figure 1. Trend of publications on Circular Economy (Title, Abstract, Keywords). Source: Scopus. Date: 01/11/2020.

Despite this broad interest, a widely accepted classification of Circular Economy topics is still missing in literature. The classifications differ one from another in terms of number of topics, depth of study and topics considered[10]. In my thesis, I highlighted **7** macro topics common to the different classifications. These 7 classes include all the sub-topics covered by the literature classification; they are: new products, efficiency, resources, industries, waste management and sustainability.

Regarding green skills, on the basis of scientific articles review, I can say that green jobs require: a good level of basic requirements, the ability to solve problems and management skills[6]. However, given the existence of different CE sectors, there already are many green profiles. There cannot be a single course of study or training: different professional figures with a high level of specialization are needed, based on the occupation they cover. These skills will be acquired with experience and continuous on-the-job training[6].

[10] Mahanty, S., Boons, F., Handl, J., Batista-Navarro, R. (2019). Studying the Evolution of the 'Circular Economy' Concept Using Topic Modelling. Lecture Notes in Computer Science. 11872 LNCS, pp. 259-270

3. Methodology

In order to identify CE macro-areas of research and the skills and job profiles needed for this paradigm, I set five macro-tasks:

- **papers retrieval**: the aim of this step is identifying and collecting scientific papers dealing with Circular Economy. This task is fundamental for the whole outcome of the analysis; for this reason, I made a great effort to build the query on Scopus[11] (a database of abstracts and citations launched in 2004 by Elsevier);
- **topics identification**: I used the Topic Modeling algorithm, a Text Mining tool, to identify CE topics and I carried out three steps:
    1. pre-processing of the data: for the success of Topic Modeling it is important that in the documents of interest the words considered are as relevant as possible to the topic studied;
    2. Topic Modeling, with the use of the LDA algorithm (Latent Dirichlet Allocation). The LDA is based on the intuition of Blei et al. (2003)[12] that each document can be seen as a mixture of topics and the topics can be seen as a mixture of words. This algorithm returns for each paper and for each word the probability of belonging to the different topics: the probabilities associated with documents are called *gamma* (γ), those associated with words *beta* (β);
    3. topics labels determination, since Topic Modeling did not provide the names of the topics obtained, I carried out analyses aimed at determining them.
- **skills recognition**: in which I automatically extracted the skills from the abstracts of CE scientific publications and analysed them. The skills extraction is performed with the Nominal Entity Recognition (NER) tools, a branch of Text Mining. In particular, I used the Gazetteer Based method, which consists in the projection of a list of words (in this case the skills) in a *corpus*; if an expression present on the start list is found, it will be extracted. As a list of starting skills, I used the most important public taxonomies of occupations and skills: ESCO[13] (European Skills, Competences,

[11] https://www.scopus.com/standard/marketing.uri

[12] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

[13] https://ec.europa.eu/esco

Qualifications and Occupations) and O*NET[14] (Occupational Information Network). The first is provided by the European commission and it contains skills, competences, qualifications and occupations relevant for the EU, while United States Department of Labor designed the second.

Once I extracted all the skills from the abstracts of CE scientific publications, I analysed them in order to keep only those that actually represent skills. A word can have different meanings, for example ESCO considers Chinese a skill (language knowledge), but in publications it was used as an adjective.

- **skill-topics association**: the objective of this step was to associate the skills with the identified CE topics. I made the association on the basis of the papers probabilities to belong to the different topics (*gamma* values). Once this was done, I carried out a skill analysis divided by topic;
- **green job profiles mapping**: to arrive at the mapping of green job profiles, I carried out three steps. Firstly, I extracted from ESCO, using the Nominal Entity Recognition (NER) tools, the CE job profiles. To do this, I used the information related to the different topics (common words and skills) and the ESCO job profiles description. Secondly, I carried out an analysis of the correlations between job profile and topic. Finally, I ran a profiles classification.

4. Analysis and results

In the following sub-paragraphs, I briefly reported the results achieved in each step of the methodology.

4.1 Papers retrieval

I built a CE centered query to create a scientific publications database; I executed this query on Scopus[11]. For the success of the analyses, I made sure that the database precision (the fraction of CE papers among the recovered ones[15]) was the maximum possible. In order to do this, I set in the query that scientific publications must contain (in the title or in the abstract or between the keywords) both “Circular Economy” and another CE keyword. This query generated on Scopus **6.461** articles.

[14] https://www.onetonline.org/

[15] https://it.wikipedia.org

I calculated the precision of the database analyzing a sample of 50 articles for each year in which publications existed; for the years when the publications were less than 50, I took them all. The sample thus obtained consisted of 712 papers. The precision was **99.72 %**; only two of the verified scientific publications were not CE related.

4.2 Topics identification

After having carried out the data pre-processing and applied the LDA algorithm, I divided the scientific publications into **7** topics. To associate labels to the 7 topics, I analysed for each one of them the 15 articles with the highest *gamma* value (the probability of belonging to the topics). For each article I determined, by choosing the most appropriate among the keywords, the main theme. Then, I associated with each topic the most frequent keyword among the 15 articles analysed. In relation to the chosen keywords, I matched with the different topics one of the literature classes found in state of art review. In *Table 1* I report for each topic the most frequent keyword in the 15 articles and the respective literature class. Only in one case I did not find an exact match.

Topic	Most frequent keyword	Literature class
1	Eco-innovation	New products
2	Eco-efficiency Indicators	Efficiency
3	Material scarcity and product life cycle	Resources
4	Sustainable management and business model	Industries
5	Environmental engineering	-
6	Waste management	Waste management
7	Sustainable development	Sustainability
-	-	Pollution

Table 1. Number of scientific publications divided by topic. Source: Scopus. Date: 01/11/2020.

The class found in the literature that does not have an exact match is "Pollution"; the topic which remains excluded from this association is "Environmental engineering". I can say that while the classifications I found in literature focus on problems, scientific publications tend rather to see resolutions[7]. For the 6 topics with a match, I gave the literature classes labels (for instance, “New product”, “Efficiency”, etc.); for the seventh, I named the label "Environmental engineering".

I validated the choice of topics labels through the analysis of: the most frequent words, the words with the highest *beta* and the publications of the most frequent authors. *Figure 2* shows the ten words more frequent for the first 3 topics; for example, topic two, (that gets the label "Efficiency") has, among its ten words most frequent, both “indicator” and “eco-efficiency”.

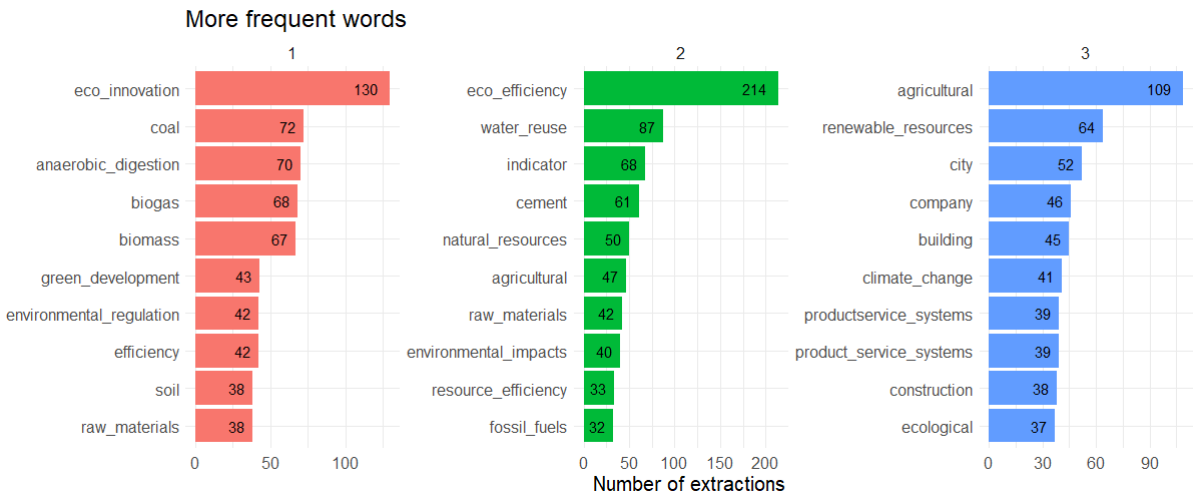

Figure 2. The 10 words most frequent for the first three topic: (1) New product, (2) Efficiency and (3) Resources.

4.3 Skills recognition

The hypothesis behind the operation of skills extraction is: if a skill is cited in a CE scientific publication, then this skill concerns the topic.

The technique used for this extraction is the Gazetteer Based one. I projected both the ESCO[13] and O*NET[14] skill lists on the abstracts of scientific publications. From the list of ESCO skills, consisting of 97.204 expressions, I extracted **1.044** different skills; from O*NET list, made up of 119 skills, only **33**. Since the skills extracted from ESCO contained all those extracted from O*NET, I carried out the analysis phases only on the skills extracted from the first.

I cleaned the list of skills obtained to keep only the expressions that really represented a skill in the sentences. Then, I calculated the precision of a 200 sentences random sample; the precision was **91 %.**

4.4 Skills-topics association

The skills association to the 7 CE topics was fundamental for the job profiles extraction. Using Topic Modeling, I associated to each paper the probabilities of belonging to the topics (the *gamma* values). For each paper, I first chose the topic with greater *gamma* probability; then, I associated all skills (extracted from the abstracts) to that topic. In *Table 2*, I report three examples of skills for each topic. The skills in *Table 2* were among those extracted several times, the most relevant for each of the 7 topics.

New products	Efficiency	Resources	Industries	Environmental Engineering	Waste Management	Sustainability
energy conservation	risk management	logistics management	fertilizers	logistics management	environmental legislation	environmental sustainability
Petroleum	economics	fossil fuel	food manufacturing	coordination	hazardous waste	ecosystems
production process	logistics management	solar energy	electricity	innovative thinking	ethanol production	supply chain management

Table 2. Examples of skills extracted and associated with each topic.

From these few examples, it was possible to see that:

- in some case, the examples were a further validation of the labels assigned to the topics. For the topic "New products" I found “petroleum”, a resource that is running out and, for this reason, the products as we know will change; another example could be for the topic "Resources", for which I found “solar energy”;
- the examples confirmed what was in the state of art study regarding skills (namely, that among the green skills there are solving problems and management[6]). Among these examples there are: “logistics management”, “supply chain management”, “risk management” and “innovative thinking”;
- the examples confirmed also CE transversality (established in the state of art). The skills found belong to different sectors, such as “food manufacturing”, “electronic products”, “ethanol production” and “livestock farming system”.

4.5 Green job profile mapping

The ESCO database lists the existing job profiles in Europe; for each profile, it gives a brief description and a set of essential skills. I identified CE job profiles using the information provided by ESCO. I searched, in relation to each CE topic, the ESCO job profiles that include:

- in their description at least one of the topic most frequent words;
- in their essential skills at least one of the topic most extracted skills.

I extracted **85** different job profiles; the same profile may have been extracted from different topics. The graph in *Figure 3* reports each job profile extracted and the 7 CE topics. The graph is bipartite, namely the nodes represent both topics and job profiles; the nodes relating to the 7 topics stand out because they have larger and bolder writings. The arcs connect the profile nodes to the topics from which they were extracted.

Figure 3. Graph relating to the connection between extracted professional profiles and CE topics.

The first thing I can observe about the graph is the fact that the job profiles were consistent with the topics from which they were extracted. For instance, profiles relating to waste (“hazardous waste inspector”, “landfill supervisor”, “solid waste operator”, etc.) are linked to "Waste management"; profiles relating to environmental (“environmental education officer”, “environmental programme coordinator”, “environmental expert”, etc.) are linked to "Environmental engineering".

The graph also assigns to the most connected topics and job profiles the same colour, therefore it is possible to see three different areas (orange, blue and green). The topics that are correlated for each area and the shared profiles are consistent one with the other:

1. in the orange area there are the topics "Waste Management”, “Industries” and “Sustainability”. Waste produced by industries is a large part of sustainability issue. For instance, the “hazardous waste inspector” oversees industrial sites to ensure they adhere to waste disposal legislation and to sustainability regulations[13];
2. in the blue area there are the topics “New product” and “Resources”. Resources, whether they are running out or renewable, are strictly correlated with the design of new products that are respectful of the planet. For instance, the "renewable energy engineer" researches alternative energy resources, in order to design new systems for renewable energy production[13];
3. in the green area there are the topics “Environmental Engineering” and “Efficiency”. The environmental engineering is the application of engineering principles to

improve and maintain the environment[16]; this kind of engineers cannot set aside the efficiency in the resources, products and systems use.

Finally, I carried out a classification of the profiles obtained. First of all, for each professional figure, I searched for the corresponding figure on O*NET. I did it because O*NET labels the profiles as green or not. After further analysis, I divided the 85 profiles (extracted as described above) into 4 classes:

- 69 profiles considered green by O*NET (agricultural engineer);
- 4 profiles not considered green by O*NET, but that I can consider green (aquaculture site supervisor);
- 8 profiles not belonging to CE areas, which will remain in support of the green ones (civil engineering technician);
- 4 profiles not belonging to CE areas (mine production manager).

I can formulate two hypothesis about the last profile class: one is that the four profiles are outliers of the analysis; the second is that scientific publications can also talk about the professional figures that could disappear or be replaced with CE applications.

5. Conclusions and future works

In this thesis I show an approach to find CE sub-topics and to map green job profiles.

The topics classification can be considered satisfactory because the topics are consistent with the classifications found in the literature study for what concerns meaning, number and grain. Furthermore, the approach aims to collect professional profiles related to CE paradigm and to link them to specific CE area.

Potential stakeholders of my work are summarized in *Table 3*, together with the hypothesized benefits that they can obtain from the two results of this thesis.

	Stakeholders			
	Public Administration	**Businesses**	**Researchers**	**Students**
CE Topics	Knowledge of the general topic trend	Knowledge of the general topic trend	Scientific consciousness on the domain	Training
Green Job Profiles	Industries that can receive green bonuses hiring these workers	Necessary professional figures in the company	Professional profiles to be trained	Choice of the training path to follow

Table 3. Examples of utility that four different stakeholders can draw from the results achieved.

Further steps of my work may be: (i) the elaboration of training courses for CE professional figures; (ii) the formalization of the process of determining the profiles as green or not, in order to add this information to the ESCO database.

[16] https://www.aaees.org/careers/