

UNDERSTANDING THE BCORP PHENOMENA: USING WEBSITES TO CLASSIFY COMPANIES

BY MIKAËL HÉROUX-VAILLANCOURT

P4IE PRE-CONFERENCE 2022



POLYTECHNIQUE
MONTRÉAL

WORLD-CLASS
ENGINEERING

AGENDA

Context

Data acquisition and Treatment

Clustering results

Discussion

CONTEXT | MY RESEARCH

- Member of the Canada Research Chair in Management and Economics of Innovation and 4POINT0
- Research of new innovation indicators from the corpus of companies' websites.
- I am particularly interested in determining whether a signal observed in the corpus of a set of companies does indeed correspond to a dimension related to innovation and whether this observation does indeed have a valid epistemic value.
- Can we use the content of company websites as complementary or substitute data sources to the data obtained via questionnaires?

Héroux-Vaillancourt, M., Beaudry, C., Rietsch, C., 2020. Using web content analysis to create innovation indicators—What do we really measure? Quantitative Science Studies 1, 1601–1637. https://doi.org/10.1162/qss_a_00086

CONTEXT | TRADITIONAL DATA SOURCES

1. Questionnaire-based survey

- Precise information vs cost due to low response rate (*Sauermann, 2013*)
- Several methodologique biases (Acquiescence bias (*Watson, 1992*), demand characteristics biases (*Orne, 1962*), extreme responding (*Furnham, 1986*), self-reporting biases also called social desirability biases (*Nederhof, 1985*), selection biases (*Heckman, 1979*), non-response biases (*Deming, 1990*))

2. National statistics databases (Statistic Canada, USPTO)

- Specific indicators, lack of context (p. ex. \$ de R-D, # of patents)
- Difficulty to access databases

CONTEXT | POSSIBILITY TO USE ENTERPRISES' WEBSITES INFORMATION AS A RELEVANT DATA SOURCE?

Web content analysis of tech enterprises' websites

Advantages

1. Access for everyone
2. Lots of information available
3. Frequent update
4. Available at any time

Drawbacks

1. Unstructured data
2. Significant differences between sites
3. Self-reporting bias

Web-mining indicators could represent new information that was not captured through classical methodologies and could possibly be complementary to traditional data sources (Gök et al., 2014; Héroux-Vaillancourt et al., 2020)

Context | Culture and performance



All practices that are considered important for the enterprise
It comprises activities that can be considered to be performed

CONTEXT | REDEFINING PERFORMANCE IN A PERFECT STORM OF GLOBAL CHALLENGES

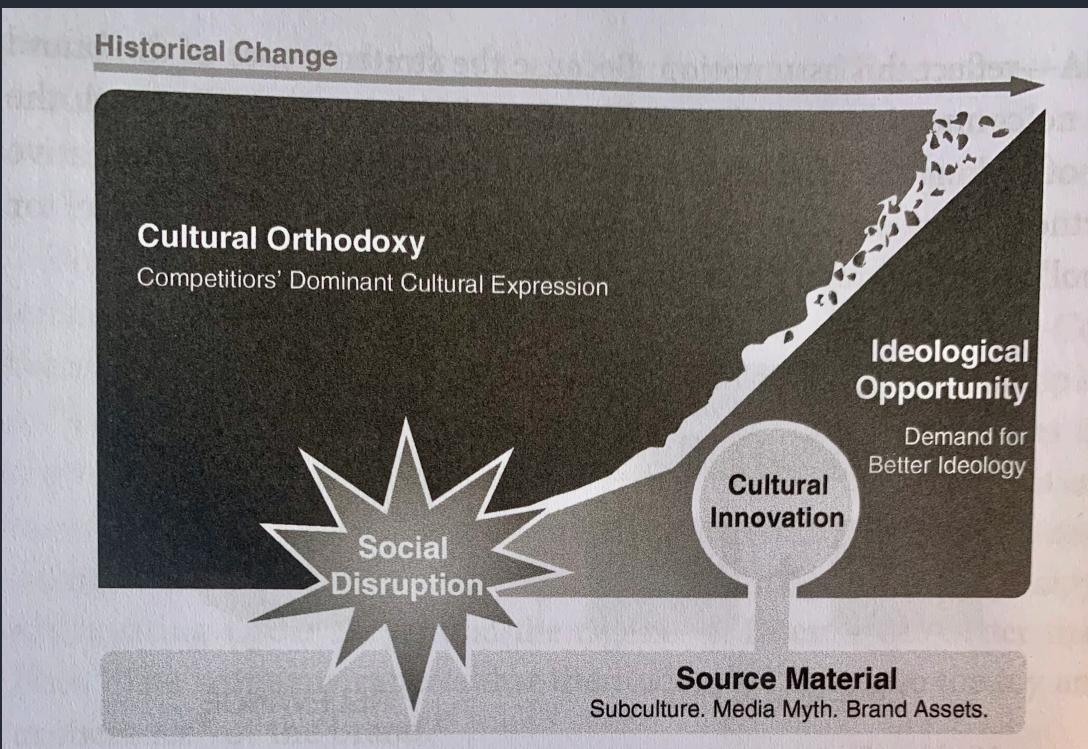
- Economy
 - High inflation , raise of interest rates, market volatility, COVID relief measures
- Environment
 - Climate change, biodiversity loss and global pollution
- Society
 - Gender equity (#metoo, pay equity)
 - Identity issues (BLM, First Nations relations, decolonial movements)
 - Polarisation of opinions, loss of trust in experts and political radicalisation
 - Wealth gap
- Individual level
 - Record increase in depression, suicide and loss of meaning in life in developed countries (See the series "Awakening from de Meaning Crisis" by John Vervaeke of the University of Toronto)

How does a 21 century company "perform" in regards to these massive challenges?

CONTEXT | CORPORATE SOCIAL RESPONSABILITY (CSR)

- “*... there is one and only one social responsibility of business to use its resources and engage in activities designed to increase its profits so long as it stays in the rules of the game, which is to say, engages in open and free competition, without deception or fraud.*” – Milton Friedman, 1970

CONTEXT | CULTURAL INNOVATION



Holt, D., & Cameron, D. (2010, p.12)

There is an ideological opportunity for purpose-driven companies

CONTEXT | CSR AND INNOVATION

- There is emerging literature between adopting CSR strategies and innovation performance
 - Environment and innovative performance (Borger & Kruglianskas, 2006)
 - Service and product innovation (Zhou et al., 2020)
 - Firm's responsible performance, leads to innovation and strong social benefits (Wagner, 2010; Martinez-Conesa et al., 2017)

CONTEXT | B CORP

- Business model that shifts priorities from shareholder's perspective to stakeholder's perspective
- The B Corporation (B Corp) audit and certification can be seen as a signal for CSR business model innovation ([Moroz & Gamble, 2021](#))
- Certification data is frequently up to date, transparent on [data.world](#) and free to use
 - 8799 certifications from 5631 companies across 86 countries since 2006
 - Data includes among others certification scores, company's website, industry, size and location
- Can we use the website of these companies to better understand the context of the B Corp phenomenon?



DATA ACQUISITION AND TREATMENT

DATA ACQUISITION

Extract all the homepage
of certified US and
Canadian companies from
the database from 2007 to
2022 (1741)

- Used [an open source Wayback machine scraper](#) to scrape all homepages for Canadian and US URLs

Other selection criteria
(1256)

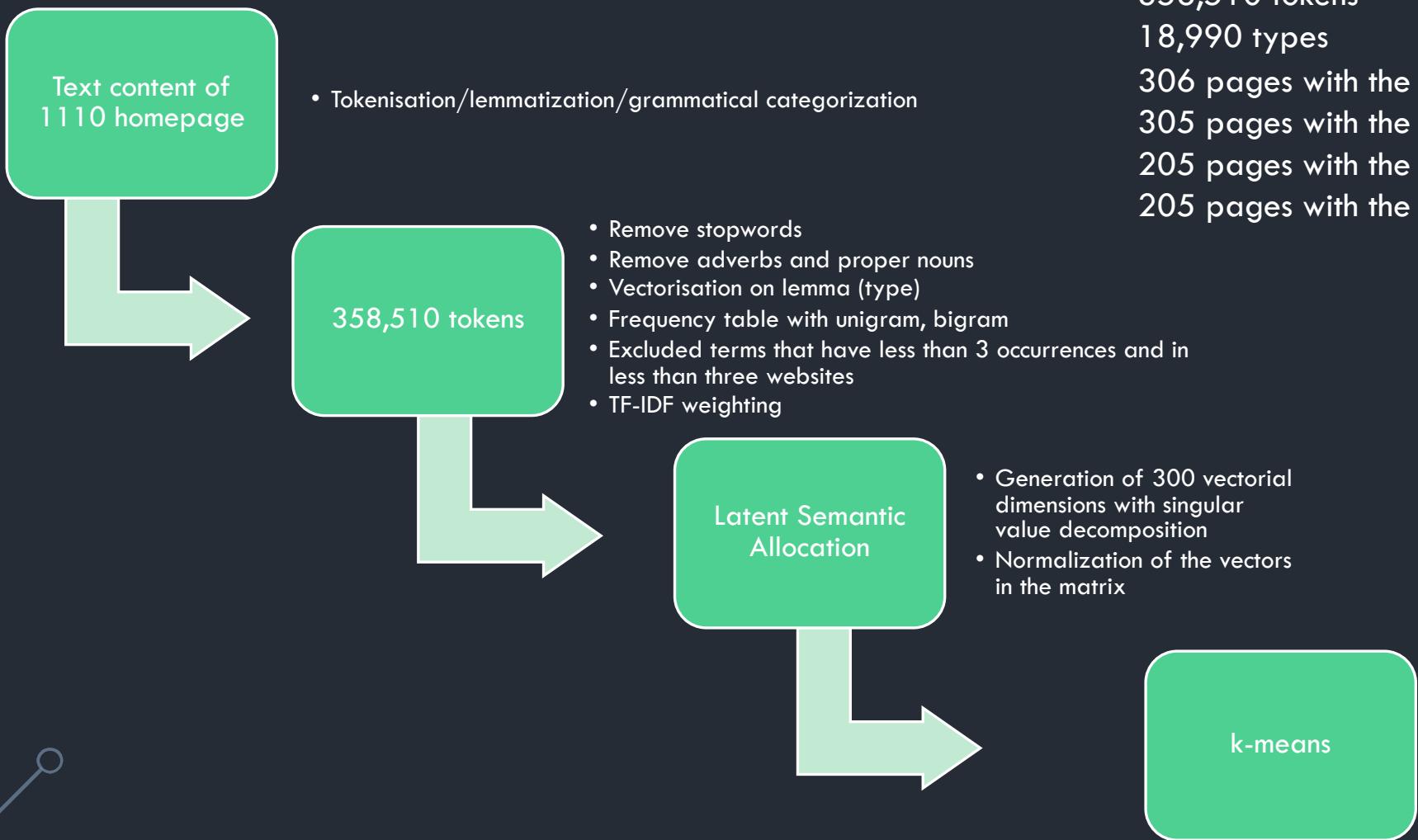
- Select only English pages
- Select the most recent audit
- Remove duplicates

Final dataset (1110)

B CORP SAMPLING

- For the context of this research, we selected all American and Canadian companies and scraped the homepage the most recent year of each B Corp audit:
 - 1110 unique homepage content
 - Location: 82% US, 18% Canada
 - Size: 5% Big enterprises, 95% SMEs, 78% Small enterprises, 45% Really small enterprises (< 10 employees)
 - Assessment years: 5.4% 2021, 19.8% 2020, 30.5% 2019, 26% 2018, 15.9% 2017 and 2% 2016-2015
 - 13.3 % Food and Beverage, 8 % ITC, 7.65% Marketing, 5.95 % Home and personal care, 5 % Investment advisor, 4.6% Business consulting, 4.23% Apparel, 3.15%, Renewable energy, 2.97% Education, 2.47% Sustainability consulting, 42.7% among 56 other industries

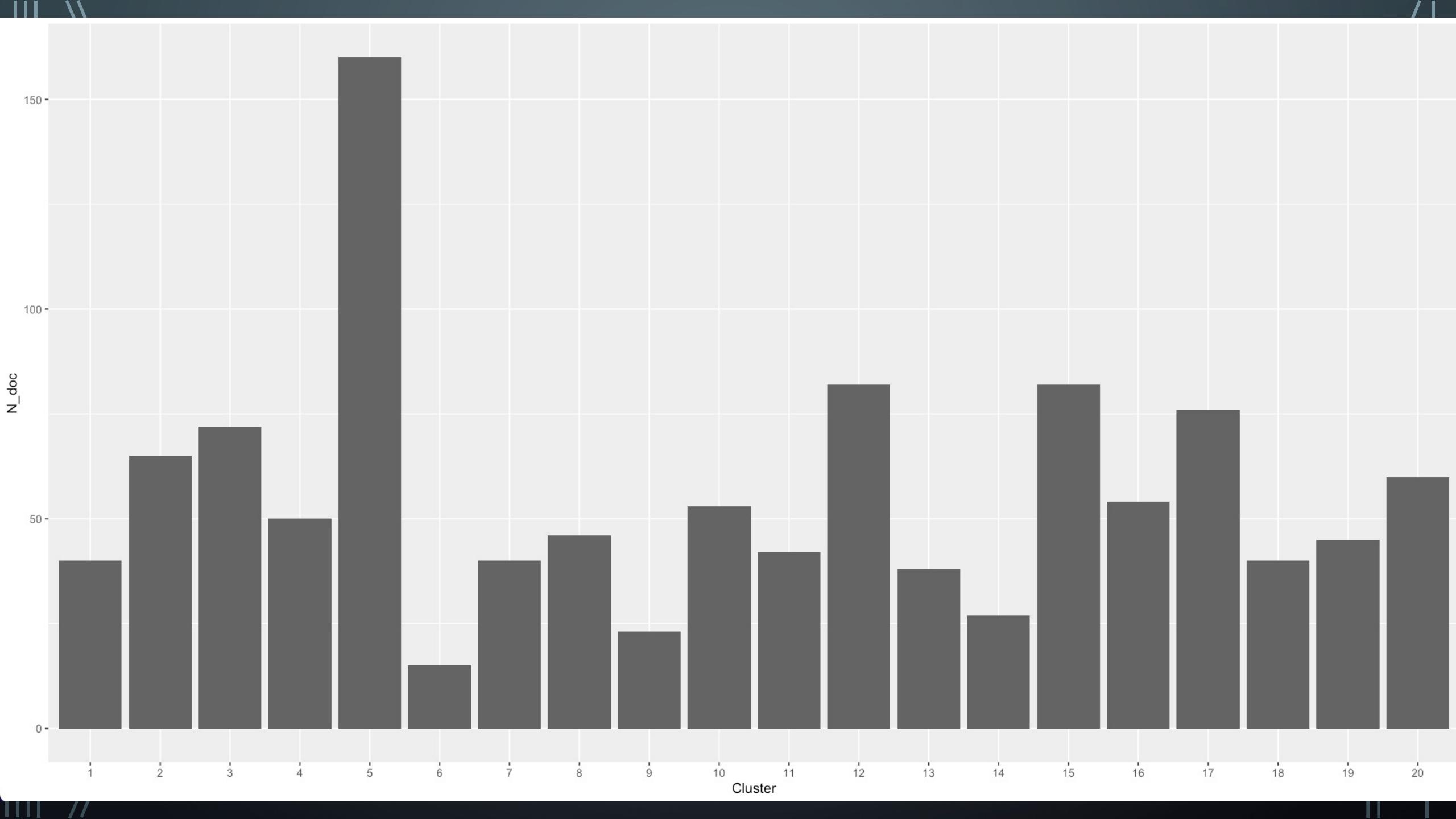
CLUSTERING PIPELINE



1110 homepages
358,510 tokens
18,990 types
306 pages with the term 'social'
305 pages with the term 'environment'
205 pages with the term 'innov'
205 pages with the term 'invest'



CLUSTERING RESULTS





CLUSTER 1

n=40

Words hinting Bcorp signals :
0/30

investment_advisory
responsibility
investment_advisor planning help life advisory_service
fiduciary service register plan provide
individual wealth advisor retirement decision
management service_offer client firm offer advice
security financial_planning register_investment offer_register





CLUSTER 2

n=65

Words hinting Bcorp
signals: 9/30



A word cloud visualization for Cluster 2, featuring words related to entrepreneurship, business, and social impact. The most prominent words are 'entrepreneur' (yellow), 'business' (dark brown), 'impact' (pink), 'company' (red), and 'solution' (purple). Other visible words include 'sustainable', 'growth', 'capital', 'people', 'social', 'startup', 'community', 'firm', 'global', 'help', 'world', 'create', 'value', 'drive', 'partner', 'ecosystem', 'organization', 'purpose', 'economic', 'venture', 'mission', 'idea', 'believe', 'build', 'positive', and 'believe'. The words are colored in various shades of green, yellow, orange, red, purple, and blue.



CLUSTER 3

n=72

Words hinting Bcorp signals:
5/30

product





CLUSTER 4

n=50

Words hinting Bcorp
signals: 0/30

A word cloud visualization showing various financial terms in different colors. The words include: deposit, banking, conversion, service, pay rate, website, cash, secure, account, loan, union, book, financial, money, time, feature, sell, cookie, insurance, information, payment, access, online, announce, start, mortgage, and customer.



CLUSTER 5

n=160

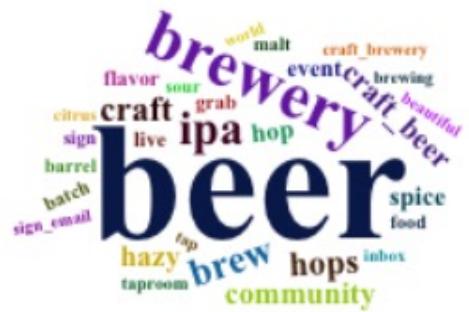
Words hinting Bcorp
signals: 2/30



CLUSTER 6

n=15

Words hinting Bcorp
signals: 1/30



CLUSTER 7

n=40

Words hinting Bcorp
signals: 5/30

student

school

education

teacher

college

water

program

library

filter

charter

community

service

public

adventure

learning

help

child

safe

degree

learn

empower

son

urban

classroom

support

graduate

public_school

CLUSTER 8

n=46

Words hinting Bcorp
signals: 13/30



CLUSTER 9

n=23

Words hinting Bcorp signals: 8/30



CLUSTER 10

n=53

Words hinting Bcorp
signals: 3/30

compassionate
therapy
mental health
family
support
life
mental program
offer
disease
wellness
provide
addiction
dental
include
treatment
employment
help
improve
provider
employee
pregnancy
health-care
healthcare
class
development team

CLUSTER 11

n=42

Words hinting Bcorp signals: 2/30





CLUSTER 12

n=82

Words hinting Bcorp
signals: 9/30

organization





CLUSTER 13

n=38

Words hinting Bcorp
signals: 4/30



CLUSTER 14

n=27

Words hinting Bcorp
signals: 9/30



CLUSTER 15

n=82

Words hinting Bcorp signals: 2/30

A word cloud centered around the word "organic". Other prominent words include "ingredient", "delicious", "food", "plant", "seafood", "planet", "farm", "cheese", "honey", "chocolate", "day", "pack", "product", "source", "hen", "taste", "eat", "free", "protein", "sugar", "dairy", "quality", "animal", "healthy", "fruit", "meal", "health", "snack", "love", and "taste". The words are in various colors and sizes, with "organic" being the largest and most central.

CLUSTER 16

n=54

Words hinting Bcorp
signals: 5/30

brand

organization business elevate
help_brands build impact
agency tell_story
production company media social partner
change video transform consumer
creative help
help_leader story tell
mean_help purpose people client
world

CLUSTER 17

n=76

Words hinting

Bcorp signals: 0/30

organization
experience
technical
provide
application
system
expertise
service
client
help
it
management
grow
device
legacy
business
firm
solution
customer
team
technology
software
company
value
build
support
project
service_business

CLUSTER 18

n=40

Words hinting Bcorp
signals: 1/30

international
digital_marketing
marketing
digital
client
design
development
help
reach
build
craft
project
business
service
strategy
marketing_agency
campaign
marketing
team
media
digital_agency
content
brand
audience
website
social
digital_strategy
seo
web_design
agency
marketing_agency

CLUSTER 19

n=45

Words hinting

Bcorp signals: 5/30

investment investor management company real social asset fund value return impact firm financial capital portfolio asset_management impact_investment business strong institutional management_firm strategy sustainability asset_investing investment_strategy investment_management

CLUSTER 20

n=60

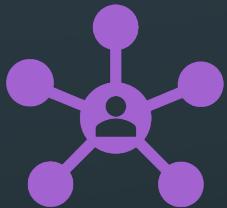
Words hinting Bcorp
signals: 0/30

A word cloud visualization for Cluster 20, showing various words related to shipping, gift-giving, and offers. The words are colored in a variety of shades, including purple, green, blue, red, and yellow, and are arranged in a roughly circular pattern. The most prominent words include 'ship', 'gift', 'free', 'discount', 'store', 'offer', 'product', 'contact', 'backpack', 'freight', 'valid', 'item', 'add', 'holiday', 'email', 'enjoy', 'make', 'date', 'stay_date', 'offer_access', 'purchase', 'code', 'give', 'stuff', 'iron', 'das', and 'offer'. The word 'ship' is at the top center, 'gift' is in the middle, 'offer' is on the left, and 'discount' is on the bottom left.

OBSERVATIONS



We can observe sectors and added context from the homepages



Most of the language that emerges in our cluster is not necessarily different from traditional business language



Some clusters really put forward Bcorp signaling such as cluster 2, 8, 9, 12 and 14

WORKSHOP

- Is there a sub-corpus that you think would be interesting to study and for which you think you would get interesting clusters? Either by applying the clustering :
 - On the content filtered from the metadata of the database such as a region or a sector
 - On all the words surrounding a keyword such as 'innov' or 'environment' etc

REMINDER OF INTERESTING METADATA TO FILTER

- For the context of this research, we selected all American and Canadian companies and scraped the homepage the most recent year of each B Corp audit:
 - 1110 unique homepage content
 - Location: 82% US, 18% Canada
 - Size: 5% Big enterprises, 95% SMEs, 78% Small enterprises, 45% Really small enterprises (< 10 employees)
 - Assessment years: 5.4% 2021, 19.8% 2020, 30.5% 2019, 26% 2018, 15.9% 2017 and 2% 2016-2015
 - 13.3 % Food and Beverage, 8 % ITC, 7.65% Marketing, 5.95 % Home and personal care, 5 % Investment advisor, 4.6% Business consulting, 4.23% Apparel, 3.15%, Renewable energy, 2.97% Education, 2.47% Sustainability consulting, 42.7% among 56 other industries

FUTURE RESEARCH



THIS WAS JUST A FIRST STEP FOR THE PURPOSE OF THIS AFTERNOON WORKSHOP



A NEXT STEP WOULD BE TO GET MORE INFORMATION ABOUT OUR CLUSTERS FROM OUR METADATA AND SEE FOR INSTANCE IF SOME CLUSTERS HAVE BETTER SCORES



TOPIC MODELING WITH LATENT DISTRICT ALLOCATION PER SECTOR WOULD ALSO BE INTERESTING TO SEE



THANK YOU!

CONTACT:
mikael.heroux-vaillancourt@polymtl.ca