I’d like to talk about one aspect of the search engine and the web ecosystem: web content production. In the first part, I’ll explain the reason why we need to know more about it. Second, I’ll introduce my method: analyzing two domestic search engines in South Korea and the Czech Republic, simultaneously. In the last part, I’ll discuss two cases by drawing from an interpretation of Marxist “primitive accumulation” from the perspective of critique of the settler colonialism.

I. Motivation: The significance of online content production for web search studies

We want search results’ to be fair and equitable because we have become concerned about search engines’ influence in shaping collective understanding of our worlds. One of the major critique of search engines (Pasquale, 2015; O’Neil, 2016; Mulligan & Griffin, 2018) is Safiya Umoja Noble’s book, *Algorithms of Oppression*. In her black feminist theorization of web search, she pointed out many problems in the representation of the people of color, especially women of color, in racist and sexist ways. In her conclusion, she argued for “public noncommercial search”: “what is needed is a decoupling of advertising and commercial interests from the ability to access high-quality information on the Internet” (Noble, 2018; 179). One of the possible ways to decouple these two, she proposed, is to visualize the information and to partition them by the sources: “[...] in a visual rainbow of color that symbolized a controlled set of categories, [...] red was pornographic, green was business or commerce related, [...] and orange was entertainment, and so forth.” In this kind of scenario, we could see “the entire indexable web and click on the colors we are interested in and go deeply into the shades we want to see” (180).

For more equitable and fairer search results to be realized, I believe that we need *not just* better search engine and control of interface *but also* web documents and content that can represent our society in nondiscriminatory ways. In other words, if “the entire indexable web” continue to reflect our world in biased manner, the extent to which search engine result get fairer
will be limited. Thus, if we want unbiased web search, we need to know more about how individuals, noncommercial organizations, and institutions as well as companies participate in creating web content. And because it would be impossible to navigate the World Wide Web without search engine, we need to understand web content creation in relation to search engine.

II. Method: inter-referencing South Korean and Czech search engines

Then, how can we study these aspects of the relationship between web content producers and search engine? It is no surprise that in the first place we would look at the relationship between Google and the English web content creators. It seems natural because almost all countries’ people use Google for their web search and thus it seems quite typical and universal phenomena to make use of Google. However, it runs the risk of taking granted a lot of socio-economic contingencies in content creation, which have shaped and have been shaped by the Internet, technology companies, and users. Thus, we need to identify those contingencies that appear inevitable but have been coproduced with new media technologies and content creation practices.

To study these taken-for-granted aspects of the relation between search and web documents creation, I propose to pay attention to other search engines outside the U.S. As of 2019, there are four non-U.S. search engines used by their own national users more than, or as much as, Google: Baidu in China, Yandex in Russia, Naver in South Korea, and Seznam.cz in the Czech Republic (Figure 1). The first three search engines have occupied the dominant market share in their national markets, and the last one had had the most portion of web search traffic until the middle 2010s and now is the second to Google in their domestic market. Therefore, studying the relationships between these non-U.S. search engines and their non-English content producers in those four cases would provide proper opportunities to study much wider, richer, and nonconformist narratives.
My attention to these other countries’ search engines except the United States’ Google is based on the postcolonial understanding of Eurocentricism. In his book, *Provincializing Europe*, Dipesh Chakrabarty problematized “the universalism of the nation-state as the most desirable form of political community” (2000/2008, 41). To make it short, to provincialize Europe means to decenter liberal political subjects in Western European countries as role models who are considered as having reached the modern state, on the one hand, and to de-peripheralize the third world people, like the peasants in Bengal, India, who are regarded as being transitional status from premodern to modern, on the other.¹ This understanding implies that if we want to decenter the hegemonic entities, we need to leverage other entities that have been regarded as the peripheral.

Among those four local search engines and content creators using the corresponding languages, I prefer South Korean and Czech cases to the Chinese and Russian ones as research subjects. It is because of the following three reasons. The first one is accessibility. As a South Korean native, I am more aware of the South Korean’s Naver than the Chinese and Russian ones.

---

¹ I have to note that one woman scholar, Professor C. Christine Fair at Georgetown University has said since 2017 that he had been involved with sexual harassment to her in 2004 when she was doctoral student in the Department of South Asian Languages and Civilizations at the University of Chicago. Then, can I continue to cite him? Or do I have to decline to do? I have no clear answer. But if we love theories and the theorists would benefit from our elaborations based on their theories, we might have to reflect whom to cite.
Second, it is because the analysis of the Chinese and Russian cases would be far politically-charged. These cases come with troublesome issues like information control by the oppressive state. Even though these issues are also serious human rights issue, the expansive consideration might distract us from studying the relationship between search engines and web document creations.

The third reason to take South Korea and Czech Republic case is to aim at making each case to be a reference point to each other. This idea is inspired by another postcolonial thinker and inter-Asian cultural scholar, CHEN Kuan-Hsing’s book, Asia as method (2010). He was originally motivated by a Japanese historian, MIZOGUCHI Yuzō’s book, China as Method (1996 [1989]). According to Chen, Mizoguchi interpreted that the Japanese modern intellectual tradition since Meiji Restoration in 1868 had tried to study ‘Asia’ as an end through the world (i.e. Europe) as means. But Mizoguchi proposed to reverse this case-method relationship and then “to conceive of the world as an end” via “China as method (or means).” Extending Mizoguchi’s proposition, Chen argued that ‘Asia’ can be conceived as method and that ‘Asia as method’ can be multiplied: “Asia can be a synonym for China, Malaysia, India, Sri Lanka, and Indonesia; or Seoul, Taipei, and Bangalore; or the third world” (254). Moreover, Asian countries can refer not necessarily to the U.S., but to each other, when they try to know themselves. He called this reference to each other, “constant inter-referencing and the dialectic of comparison” (252). When we take multiplied perspectives Chen proposed, South Korean case will be a reference point to study Czech case and at the same time Czech case will be a reference point to study South Korean case.

To sum up, I’ll take South Korean and Czech search engines and their local web document creation practices as a method to understand the relationship between Google and the English web creation as a case.

2 Strictly speaking, the Chen’s definition of ‘Asia’ is not applicable to Czech Republic in that Czech Republic is not even what was the last called, “the third world.” Rather, it is sort of one of the post Eastern Bloc countries, having gone through "transition" "on its western periphery” (Bittner, 2017). If we could call Asia as ‘non-Western,’ then we might call Central and Eastern Europe (CEE) countries as ‘barely-Western,’ if I dare to. However, the articulation, understanding, and conceptualization of the CEE in STS seem not so prominent. In this way, this undercategorized region’s and the corresponding countries’ experiences of technoscience need be acknowledged and accounted for, with sensibility in mind. Nevertheless, we can find some exceptional STS works regarding the Czech Republic such as: the assessment of STS and S&T policy scholarship in CEE in the middle 1990s (Balázs, et al., 1995); internationality building in “laboratories located in the geopolitical East of Europe” (Lorenz-Meyer, 2012), critique of the conceptual "Immutable Mobiles" based on the analysis of "the social sciences in non-Anglophone countries" (Stöckelová, 2012); user and activism studies in Czech wireless community (Söderberg, 2010; 2011; 2014); pureness of facts and the political relevance of expertise in democratic decision making (Konopásek, et al., 2008); and civic epistemology in the Czech Republic (Stöckelová, 2009). In addition, some scholars have worked about communist Czechoslovakia’s science and technology: history of computing in Czechoslovakia (Paju & Durnová, 2009; Durnová, 2010; Durnová, 2014); “technology transfer during the cold war” (Freeze, 2007); and indie game cultures and activism (Švelch, 2018).

3 My takeaway is also indebted to contemporary East Asian Science and Technology Studies (EASTS) communities who have discussed and debated East Asian scholarships’ possible contributions to the Science and Technology
III. Theory: the reinterpretation of Marxist “primitive accumulation” from the critique of settler colonialism

If we take this methodological approach, then which specific new findings will be discovered? Even though I am still at the initial stage of this research project, one of my first impressions is these search engines’ ambivalence in approaching their own technological prowess, or “heteroglossia,” the quality of a text that speaks with many voices at once” (Bakhtin, 1982; quoted in Seaver, 2017). For example, one of Naver’s official blogs, “Naver Search and Tech,” has published many so-called ‘tech-blog’ style entries, typical one of which is about Translation of Natural Language Query Into Keyword Query Using a Recurrent neural network.

However, in his interview in 2017, the Director (Leader) of Search, KIM Kwang-Hyun, attributed Naver search’s competency not only to its algorithms but also to its exclusive in-house web resources. When an interviewer asked “Naver has obtained a lot of in-house databases. What do you think its pros and cons? What’s Naver’s strategy in dealing with external websites and social media,” he answered, “Ultimately, Naver is a search service powered by in-house databases. From this view, one of the best examples in Naver’s advantages can be sort of restaurant review, which cannot be easily found in Google [Korea] search. […] If users want to find vivid reviews, they are likely to choose Naver. [However,] Google search result shows websites too much, giving dry impressions.”

In a similar vein, Seznam.cz shows a proud history of its technology. In its promotion flyer in 2018, it says: “Seznam.cz is the only one in Europe to resist the global player Google and has a considerable share of almost half the online search field [in the Czech market]. What makes Seznam.cz unique is that unlike many other “local” search engines in Europe it has its own technology for full-text searches. As early as in 1997 it launched Kompas, its own full-text search engine.” But this narrative shows partial trajectory of technology. According to the Wikipedia, “The Kompas was subsequently replaced by an outsourced service from Empyreum,” which was replaced in 2003 with global partners, Google and Jyxo.cz, which was replaced in 2005 with Seznam’s own search engine technology, which is replaced in 2009 with Seznam’s new full-text search engine that has been running since then. Furthermore, Seznam had not owned its in-house image and video search capabilities until 2015. Before the moment, it had outsourced image search to Swedish company picSearch, and video search to a Russian search engine, Yandex. Instead, one of its fundamental competencies has come from its ownership of more than 20 web portals (Figure 2).

Studies. Among some prominent agenda in EASTS communities are “Asia as Method” and “Provincializing STS.” I hope to my methodological articulation will keep benefitting from those conversations and ultimately benefitting those dialogues back.
Currently, Seznam.cz is the clear leader on the Czech market, and Internet No.1, which is proven by its financial results for the year 2017. In 2017, Seznam.cz ended with an operating profit of CZK 1.1 billion (€43.3 million), with a turnover of 3.7 billion CZK (€145.5 million). As of today, Seznam.cz operates more than 20 of its own portals, with the real number of visits exceeding two million users per day.

**Figure 2.** The 20 most visited sites on the Czech internet, among which the top 1, 2, 4, 5, 6, 8, 10 and 11 sites belong to Seznam.cz (highlighted bars in red). Above the 20 bars, this promotional flyer texts for its search advertise clients said “Seznam.cz operat..."
Both cases of Naver and Seznam show that one of the significant factors in their search competencies are their own content that they created or helped create. This unexpected key factor of two search engines invite us to rethink about the relationship between content creation and search engine. We have been told that Google’s great success over the world is attributed to its technical excellence. When web portal and web directory companies hired human editors to find and organize web documents in the late 1990s, Google tried to automate the way to find better documents to users’ queries. Google’s reliance on algorithms, 24/7 automated web bots, and computing infrastructures rather than human labor has made the company itself easily scale up its search capability.

But in this narrative, we have taken for granted the existence of a huge amount of web documents. The cases of Naver and Seznam demonstrate that web documents are not given (Figure 3). For example, Naver has consistently helped create web documents in two ways: on the one hand, by providing users with various web services like blog, forum, and Quora-like Q&A service, on the another, by purchasing the rights to service of the existing encyclopedia and professional knowledge databases in Korean. Based on these in-house databases, Naver has prevented other search engines’ web crawlers from collecting Naver’s own databases. This walled-garden strategy has been successful, leading to Naver’s over 70% search market share (Nielsen KoreanClick, 2016).

Figure 3. Language used on the Internet (Unbabel blog 2015).
This contrast between the abundant web documents in English and the not enough web documents in Korean is resonant with the recent land-based critique to the settler colonialism. Glen Sean Coulthard, an indigenous studies scholar as well as a descendant of the Yellowknives Dene First Nation, problematized the “contemporary liberal recognition-based approaches to Indigenous self-determination in Canada” (2014: 16). According to him, Frantz Fanon, a pioneering postcolonial thinker (1968), suggested that “in contexts where colonial rule is not reproduced through force alone, the maintenance of settler-state hegemony requires the production of what he liked to call ‘colonized subjects’: namely, the production of the specific modes of colonial thought, desire, and behavior that implicitly or explicitly commit the colonized to the types of practices and subject positions that are required for their continued domination.” (16).

This mode of production of ‘colonized subjects’ and maintenance of that mode is exacerbated by the dispossession of First Nations’ lands. Coulthard argued that Marx, obsessed by European contexts and little concerned about the contexts in the ‘New Worlds,’ primarily addressed the "separation of workers from the means of production" than the colonial separation of the indigenous from their inherited land. In contrast, “the history and experience of dispossession, not proletarianization,” Coulthard observed, “has been the dominant background structure shaping the character of the historical relationship between Indigenous peoples and the Canadian state” (13). Therefore, Canadian “indigenous struggles against capitalist imperialism are best understood as struggles […] not only for land, but also deeply informed by what the land as system of reciprocal relationship (13) […] ought to teach us about living our lives in relation to one another and our surroundings in a respectful, nondominating and nonexploitative way” (60).

From this perspective of the land-based critique of settler-colonial politics, now we can identify Naver’s ambivalent situation: it has secured the Korean web not to be exploited by foreign search engines, however, it does not necessarily mean that Naver has managed the Korean web “in a respectful, nondominating and nonexploitative way.” We have to identify more subtle ways in which the digital political economy in local scale is working in layered with digital political economy orders in global scale. It would help us better understand other countries’ native people’s lives under settler-colonial norms as well as most countries’ Internet users that have no other option for search without Google.

**IV. Concluding Remarks**

I have still many things unexamined and untested: I skipped to show examples of inter-referencing between Naver and Seznam; Google and English web as case has not been analyzed; and different theoretical foundations need to be triangulated from each other. Nevertheless, I hope that the decentered understanding of the production and navigation of digital information
would bring us new ideas to address our contemporary concerns about digital identities, representations, and self-determination. Thank you.

References


Furthermore, studying these aspects of the relationship between textual web content producers and search engine can be a roundabout way to circumvent perils of rendering “algorithms as closed boxes to be opened” (Seaver, 2017). It is well known that it is so easy to be blocked or get obstructed when trying to see through algorithms and to peek what’s inside the so-called black box. Moreover, algorithms are not static artifacts, but open to consistent changes, updates, tweaks, and performance enhancements, over time (Seaver, 2013). However, this way of understanding algorithms as beyond investigations ends up at many times rendering algorithms as unfathomable mysteries and then reifying its legitimacy (Ziewitz, 2016; Ananny, M., & Crawford, K. 2018). In contrast, if we follow how search engine relate itself with web text producers and how the relationships have changed, we would comprehend more aspects of “heterogeneous and diffuse sociotechnical systems” (Seaver, 2017).

Many of us might be familiar with the impression of the Chinese and Russian internet as the oppressive control of information, self-sufficient, and confined information network. They are repeatedly criticized by Western commentators due to their lack of freedom of speech and the limited access to information. For example, the Chinese government has prevented its people from accessing U.S. internet services (like many Google services and Facebook) as well as Euro-American newspapers’ websites (such as nytimes.com and lemonde.fr). Wikipedia (n.d.) “Websites blocked in mainland China” https://en.wikipedia.org/wiki/Websites_blocked_in_mainland.China (accessed in April 10, 2019).

At the same time, one thing I would like to note is that this understanding comes from the existing framings and characterizations under Western eyes. To take another example, on the same experiment by Russian government, we can identify the differences of narratives regarding geopolitical cyber networks. While BBC did not tell much in its title, “Russia considers ‘ unplugging’ from internet,” Newsweek’s title explicitly addressed political concern, “Russia Prepares for Cyberwar by Cutting All Domestic Internet From the World Wide Web.” But South China Morning Post added some defensive details to its title, “Putin wants his own internet amid potential threat of the US isolating Russia in cyberspace.” It would be quite tricky. It might be a rabbit hole, showing the Cold-War understanding of the worlds still works.

Another advantage of inter-reference might be to render them not so much Idiosyncratic. If we would compare South Korean search engine directly with Google, or Czech search engine with Google, we might end up concluding phenomena found in this search engine and its national creators are peculiar. But if we would put together two search engines and their corresponding creators, we might identify the recurrent themes and similar patterns between two.
e.g. European modern state building.
vi e.g. Chinese people’s efforts to build democratic nation-state.

‘The world that conceives of China as method is a multiplied world, in that China is an element of its composition. In other words, Europe is also an element’ (Mizoguchi 1996 [1989], 95; quoted in Chen 2000, 252).

Naver’s official blog has published as many entries as 425 since 2010. These whole entries are divided into four categories: Service, Technology, Operation policy, and People and Culture. Among four categories, the recent emphasis is upon ‘Technology’ category. For example, service category includes the most entries as many as 306 since 2010. Typical entries under Service category are flight ticket search (https://blog.naver.com/naver_search/220442691957), search engine optimization guide for mobile search (https://blog.naver.com/naver_search/220443657639), and file search (https://blog.naver.com/naver_search/220601123439). On the other hand, Technology category started at 2016.


In other promotional interview, Seznam identified itself as having accomplished more than Baidu, Yandex or Naver: “Seznam.cz has grown to become the only latin-alphabet based full-text search service in the world that is beating Google in its own market: the Czech Republic.” (Waldo, 2014).


Novinky.cz (news since 1998), super.cz (tabloid news), seznamzpravy.cz (Seznam.cz’s own broadcasting station’s homepage since 2018), sport.cz (sports news), mapy.cz (map), prozeny.cz (a magazine for women), and stream.cz (Seznam.cz’s Internet TV station).

This philosophy becomes more obvious when we see their main page, filled with the whole bunch of news updates, magazine articles, images, videos clips, and display advertisement (Figure 2; Figure 3). This aspect is quite different from the received history of search engines and web portals in the United States. Historians of the Internet and the Web have shown that the functions of search engines and web portals were separate, then merged, and ultimately became separate again (Haigh, 2008; Van Couvering, 2008). First, around the late 1990s, there were many search engines and web portals (or web directories) and these two had their own strategies. And during the ‘dotcom bubble’ era before early 2000, search engines and portals merged other web services to increase their market values by vertically integrating various web services to gain more traffic that was expected to draw income. But between 2000 and 2002, it was convinced by many investors that viable business models to convert traffic into money was hard to create and this collective conviction led to the collapse of the bubble. This collapse made web portals give up integrating search engines and then decide to outsource search features to other expertise search providers like Google and Yahoo! It was argued that this reliance of web portals on the external search engines stemmed from the fact that the keywords advertisement of search engines, pioneered by Overture, became the only feasible business model.

Since the Canadian federal government’s “Statement of the Government of Canada on Indian Policy, or known as the “White Paper” was released in 1969, the recognition paradigm has held a hegemonic position. But Coulthard saw this recognition paradigm did not contribute to the advancement of the lives of First Nations people; rather, he argued that Canadian settler-colonialism “remains structurally oriented around achieving the same power effect it sought in the pre-1969 period: the dispossession of Indigenous peoples of their lands and self-determining authority’ (Coulthard, 2014: 25).