

Beyond CO₂ Emissions: The Overlooked Impact of Water Consumption of Information Retrieval Models

Guido Zuccon
The University of Queensland
Brisbane, Australia
g.zuccon@uq.edu.au

Harrisen Scells
Leipzig University
Leipzig, Germany
harry.scells@uni-leipzig.de

Shengyao Zhuang
The University of Queensland
Brisbane, Australia
s.zhuang@uq.edu.au

ABSTRACT

As in other fields of artificial intelligence, the information retrieval community has grown interested in investigating the power consumption associated with neural models, particularly models of search. This interest has become particularly relevant as the energy consumption of information retrieval models has risen with new neural models based on large language models, leading to an associated increase of CO₂ emissions, albeit relatively low compared to fields such as natural language processing. Consequently, researchers have started exploring the development of a green agenda for sustainable information retrieval research and operation. Previous work, however, has primarily considered energy consumption and associated CO₂ emissions alone. In this paper, we seek to draw the information retrieval community's attention to the overlooked aspect of water consumption related to these powerful models. We supplement previous energy consumption estimates with corresponding water consumption estimates, considering both off-site water consumption (required for operating and cooling energy production systems such as carbon and nuclear power plants) and on-site consumption (for cooling the data centres where models are trained and operated). By incorporating water consumption alongside energy consumption and CO₂ emissions, we offer a more comprehensive understanding of the environmental impact of information retrieval research and operation.

CCS CONCEPTS

• **Information systems** → *Retrieval efficiency*; • **Hardware** → *Impact on the environment*.

KEYWORDS

Green IR, Deep Learning, Water Consumption, IR Sustainability

ACM Reference Format:

Guido Zuccon, Harrisen Scells, and Shengyao Zhuang. 2023. Beyond CO₂ Emissions: The Overlooked Impact of Water Consumption of Information Retrieval Models. In *Proceedings of the 2023 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '23)*, July 23, 2023, Taipei, Taiwan. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3578337.3605121>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICTIR '23, July 23, 2023, Taipei, Taiwan

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0073-6/23/07...\$15.00
<https://doi.org/10.1145/3578337.3605121>

1 INTRODUCTION

Over time, information retrieval (IR) systems have increased in complexity, evolving from simple keyword-matching models based on statistics [31] to feature-derived learning to rank models [22] to the current state of retrieval pipelines that include neural models [21, 25, 37]. Neural models based on large language models have, in particular, demonstrated exceptional performance in various tasks, such as passage and document retrieval, question-answering, cross-lingual retrieval, and domain-specific search [15, 21, 37, 38, 40, 42, 43]. However, as models increase in complexity and size, so does their energy consumption [32]. Strictly associated with energy consumption is the amount of carbon dioxide (CO₂) emissions and the *water consumption* entailed by the energy production process and the cooling of the data centers in which these models are executed. These aspects have raised concerns about the environmental impact of information retrieval [32].

Previous research has started to address the energy consumption and associated CO₂ emissions aspects of IR models and research. In particular, Scells et al. [32] have reported these factors for several popular IR methods and have outlined an agenda for environmentally sustainable IR research. However, no attention has been given to the water consumption associated with IR models. Water is a vital resource in this context as it is used for the operation and cooling of energy production systems (i.e., the power plants that provide energy to the data center) and the cooling of data centers in which models are trained and operated. As water scarcity and drought periods become increasingly pressing global issues [3, 29, 33, 34, 41], it is essential to evaluate the water consumption of IR models along with their energy consumption and CO₂ emissions, to ensure sustainable research and operation.

This paper aims to raise awareness about the water consumption of powerful IR models, providing a comprehensive view of their environmental impact. First, we summarize the current state of energy consumption and carbon dioxide emissions in IR research and the green IR agenda put forward in previous work. Next, we discuss the various factors contributing to water consumption in the context of IR models, examining both off-site and on-site consumption aspects. We then move to how water consumption can be measured. Finally, we supplement existing energy consumption estimates of popular IR models with water consumption estimates, and we empirically probe factors that influence water consumption.

By incorporating water consumption alongside energy consumption and carbon dioxide emissions, this paper aims to promote a broader understanding of the environmental impact of IR research and operation. We encourage researchers and practitioners to consider these factors in their work, ultimately contributing to a more

sustainable and responsible approach to developing, researching and deploying IR models.

2 BACKGROUND AND RELATED WORK

The following gives an overview of the current state of Green AI & IR and the importance of measuring water in how it contributes to creating meaningful measures of ‘Green-ness’.

2.1 Green AI and Green IR

Concerns about energy consumption in the broader field of artificial intelligence (AI) were heightened after the study from Strubell et al. [35], which was one of the first to highlight the high energy usage and emissions produced by large language models. This study and others before it [2, 4] have fueled a growing interest in measuring the energy consumption of research in related fields such as natural language processing (NLP) and machine learning (ML). Within the field of IR, the environmental impacts of technology at scale have been a concern for at least a decade [12].

In general, there are two approaches one can take to quantifying energy and emissions: Life Cycle Assessment (LCA) [14] and power consumption measurement. The ISO standard defines LCA as the “*compilation and evaluation of the inputs, outputs and the potential environmental impacts of a product system throughout its life cycle*” [1]. Due to its complexity and the number of resources required [12], most studies choose to measure power consumption directly. Indeed, since then, there have been many efforts to measure energy and emissions for IR systems [5, 7–11]. Given the explosion in neural models for search based on transformers [13, 15, 17, 20, 23, 24, 30, 40, 43, 44], quantifying the energy and emissions of these systems is more critical than ever.

Recently, Scells et al. [32] proposed a framework for minimizing energy usage and emissions for IR with their ‘reduce, reuse, recycle’ methodology. For each concept, they outlined several ways IR practitioners can lower their energy usage: for example, one can *reduce* the number of experiments they do; *reuse* existing pre-trained models for experiments; or *recycle* existing pre-trained models that were initially trained on one task but apply them to another task. Like many other studies before it, this research focused on energy consumption. In this paper, we extend this methodology to consider not only the energy consumption (and, by extension, emissions) but also the water consumption. To the best of our knowledge, our paper marks the first study to investigate the water consumption of IR systems, and neural IR models in particular.

2.2 Water Consumption in Data Centers

Water consumption in data centers predominantly occurs due to two distinct factors: first, as an indirect consequence of generating electricity, typically through thermoelectric power sources, and second, as a direct requirement for cooling systems that help maintain the ideal operating environment. These factors are pictured in Figure 1. While water use for electricity generation is well known, its usage for data center cooling is likely less known to IR researchers.

The increase in model complexity associated with the latest advancement of AI models has seen the need for increasingly powerful servers. High-performance servers generate significant amounts of heat, which must be dissipated to maintain optimal operating

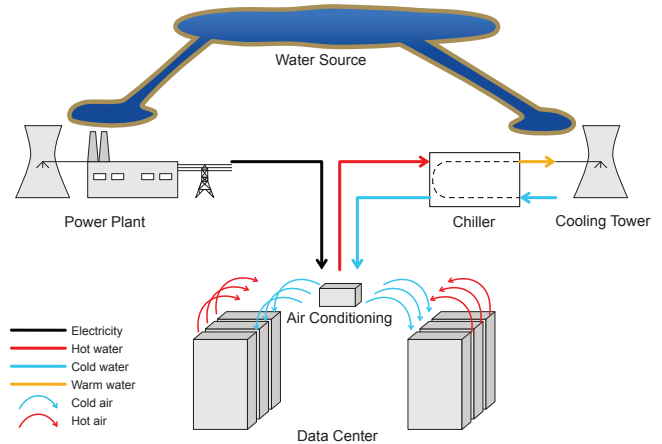


Figure 1: Water usage in a data center and a power plant.

conditions. Traditional cooling systems often rely on water, using evaporative cooling towers or chilled water systems to maintain temperatures. Consequently, the water consumption of these data centers can be significant [26]. For example, Google has reported that its data centers consumed 11.4 billion liters of water in the financial year 2017 and 15.8 billion in FY2018¹. The water used to cool data centers is often potable, thus reducing the amount of drinking water available to the population.

The type of cooling system plays a crucial role in determining water usage. For instance, air-cooled systems typically consume less water than water-cooled systems but may be less efficient at dissipating heat. Moreover, the local climate can also impact water consumption; data centers in hot and arid regions may require more water for cooling than those in cooler climates [18]. Finally, seasons and time of day also impact water consumption related to data centers cooling. The water used in data centers’ cooling towers is consumed in two ways:

- through *evaporation*, which occurs as part of the process of cooling, where hot water returning from the data center is sprayed through water distribution nozzles across a cooling fill and then collected at the bottom of the cooling tower in a cold water basis, from where water is pumped back into the chiller connected to the data center’s air conditioning; and
- through the process of *blow down*, where the water in the pipelines of the data center is flushed. This process of draining water is required to reduce salt, impurity, algae and bacteria concentration, which can cause damage to the cooling system. The higher the water quality, the less blow down of water.

Related to the blow down process is the concept of cycles of concentration S : the number of times the dissolved minerals and salts in the circulating water are concentrated compared to the concentration in the makeup water. This concept represents how efficiently the cooling tower system uses and recycles water by measuring how much water is evaporated and concentrated before discharge. A higher S value indicates that the cooling tower is more efficient in reusing water, reducing overall water consumption and

¹https://services.google.com/fh/files/misc/google_2019-environmental-report.pdf.

discharge. However, note that as the cycles of concentration increase, so does the concentration of dissolved solids, which can lead to scaling, corrosion, and other issues within the cooling system.

Water consumption related to data centers operations has attracted increasing attention from the research community [6, 18, 26], also in the context of complex and computationally-demanding AI models [19, 39]. In particular, Li et al. [19] provide a framework to estimate the water consumption of AI models and investigate the water consumption associated with the training of large language models such as LaMDA [36]. We build on the framework of Li et al. [19] to investigate the water consumption related to several common IR models.

3 WATER CONSUMPTION ANALYSIS

In the following sections, we detail a methodology for estimating the water consumption of IR models. We then complement the power and emissions estimates of several well-known IR methods from Scells et al. [32] with the water consumption of these methods. We finally break down these methods' on-site and off-site costs and analyze the influence of factors such as water quality, time of year and of the day on water consumption.

3.1 Quantifying Water Consumption in IR

We build upon the framework for measuring the water efficiency of AI models set by Li et al. [19]. In this framework, the water consumption W of a model \mathcal{M} is measured as the sum of the water consumed for the cooling of the data center (on-site water consumption, $W_{on}(\mathcal{M})$) and of the water consumed for the production of the electricity used by the data center (off-site water consumption, $W_{off}(\mathcal{M})$):

$$W(\mathcal{M}) = W_{on}(\mathcal{M}) + W_{off}(\mathcal{M}) \quad (1)$$

The on-site water consumption can then be calculated with respect to the energy $e(\mathcal{M})$ used by the IR model and the water usage effectiveness WUE_{on} of the data center. These can be parameterized by the time energy and water usage occur. This is because water consumption has daytime and season dependencies. Recall that the water is used to cool the data center. In times of the day that are hotter, e.g. in the early afternoons as opposed to the early mornings, or the hotter seasons, e.g., summer as opposed to winter, water consumption will be higher.

We account for this when computing WUE_{on} , which then is dependent from $e(t, \mathcal{M})$ and $WUE_{on}(t)$, where t represents a specific time interval (e.g., these quantities could be computed for a 15 minutes interval). Note furthermore that the power consumption of a model may not be constant across all time intervals (e.g., $e(t = i, \mathcal{M}) \neq e(t = j, \mathcal{M})$): this may be the case for example for a model that uses primarily CPU computing in a time interval and GPU computing in another time interval. Given that the water usage effectiveness WUE_{on} of the data center depends on time, running different parts of an IR model pipeline, e.g., different times of the day, may give rise to different water utilization profiles. Given this, the on-site water consumption is calculated as follows:

$$W_{on}(\mathcal{M}) = \sum_{t=1}^T e(\mathcal{M}, t) \cdot WUE_{on}(t) \quad (2)$$

Water usage effectiveness $WUE_{on}(t)$ typically depends on the cycles of concentration S associated with the blow down of water used in the cooling tower and the outside wet-bulb temperature T_w .

The cycles of concentration for a cooling tower depend on the actual cooling tower specifications and the water quality. As detailed in Section 2.2, high-quality water, i.e. with few residues and impurities, requires fewer blow downs of the cooling tower's pipelines. For example, we were able to acquire the cycles of concentration required by two cooling towers of different brands and installed in different locations of the same city in Brisbane, Australia – cooling tower A , located at a private organization, had $S_A = 2.25$, while cooling tower B , located at a public hospital, had $S_B = 1.33$.

The wet-bulb temperature is the temperature that is measured by a thermometer exposed to the air, and its bulb is covered with a wet wick, which is then exposed to moving air. As water evaporates from the wick, it cools the thermometer bulb, and the temperature reading reflects the cooling effect. Note that also T_w is dependent on time (of day and season) as it is influenced by factors such as the air temperature, humidity, air pressure, and wind speed. Thus it can be parameterized accordingly, i.e. $T_w(t)$. For example, for Brisbane where the previous two cooling towers are located, the mean annual 9am $T_w(t = 9am) = 63.5F$ (min: 53.6F, max: 71.4F), while the mean 3pm $T_w(t = 3pm) = 65.3F$ (min: 57F, max: 72.3F).

We adopt the same modeling of a cooling tower used by Li et al. [19] to compute $WUE_{on}(t)$:

$$WUE_{on}(t) = \frac{S}{S-1} \cdot \left(6 \cdot 10^{-5} \cdot T_w(t)^3 - 0.01 \cdot T_w(t)^2 + 0.61 \cdot T_w(t) - 10.40 \right) \quad (3)$$

Next we consider how to compute the off-site water consumption $W_{off}(\mathcal{M})$. The off-site water consumption is related to the cooling of the power plant, e.g., in the case of a nuclear or coal power station, and/or the actual production of the electricity, e.g., in the case of a hydroelectric power station. The off-site water consumption could also be null for some electricity production technologies, as it is the case for example for solar power generation. Similarly to the on-site power consumption, also the calculation of $W_{off}(\mathcal{M})$ depends on the energy $e(\mathcal{M})$ used by the IR model, and the water usage effectiveness WUE_{off} – which in this case refers to the power plant that generates the electricity used by the data center. However, the energy used by the IR model needs to be regulated by the relative amount of energy used by the data center to sustain that power utilization. In other words: for every kWh of electricity used by the components of a server for computation², additional energy is used by the data center to power e.g., the storage infrastructure, the power units, the pumps and ventilators used in the cooling system, etc.. This additional consumption is captured by the Power Usage Effectiveness coefficient of the data center, PUE . As for the on-site water consumption, all these quantities can be parameterized with respect to time. Thus, off-site water consumption is calculated as:

$$W_{off}(\mathcal{M}) = \sum_{t=1}^T e(\mathcal{M}, t) \cdot PUE(t) \cdot WUE_{off}(t) \quad (4)$$

²These that can be typically measured are CPU, GPU and memory consumption.

As mentioned, $WUE_{off}(t)$ is dependent on the power plant(s) used to generate electricity – and the electricity used may be produced by a mix of fuels (e.g., nuclear, coal, solar). This is modeled as follows. Be $b_k(t)$ the amount of electricity generated using fuel type k , and $EWIF_k(t)$ be the Energy Water Intensity Factor, measured in L/kWh, for fuel type k . For example, typical values for coal are $EWIF_{coal}(t) = 1.7$, and for nuclear $EWIF_{nuclear}(t) = 2.3$ [19]. Then:

$$WUE_{off}(t) = \frac{\sum_k b_k(t) \cdot EWIF_k(t)}{\sum_k b_k(t)} \quad (5)$$

Note that typically, and also in the empirical analysis below, values of k and $EWIF_k(t)$ are not known to the researchers for small time intervals t : instead, they are more likely able to source an estimate of these values based on yearly reporting from their electricity supplier or government authorities.

3.2 Water Efficiency of Common IR Models

Table 1 reports the water consumption of common IR models computed according to Equation 1 for experiments performed on the MS MARCO-v1 dataset [27]³. Along with water consumption, we also report running time, power consumption and emissions produced – these values are sourced from Scells et al. [32]. To compute water consumption, we used the energy consumption reported by Scells et al. [32]; note however that in Equation 1 $e(\mathcal{M})$ is the Power value in the table divided by the PUE of the data center. Furthermore, we used the annual mean wet-bulb temperature at 3pm in Brisbane (65.3 F), cycles of concentration $S = 2.25$, a PUE of 1.89⁴, and a combined off-site water usage effectiveness of $WUE_{off} = 1.8$, which is representative of that in Brisbane. These results assume time and season in which models are run do not influence water consumption; we consider the impact of these factors in Section 3.2. The results show an obvious correspondence between power consumption, emissions, and water consumption: as power consumption increases so do emissions and water consumption.

Figure 2 breaks down the water consumption of the considered IR models with respect to on-site and off-site consumption. This figure shows that for each model, under the considering values of the parameters, water consumption is dominated by the need of cooling the data center. This result is highly influenced by (1) the specifications and quality of the cooling tower, which influence the second part of Equation 3, (2) the quality of the water used for cooling, which in turns influences the cycles of concentrations S , (3) and the wet-bulb temperature measured at the location of the data center, which in turns is influenced by both the local climate and the season and time of day in which computations occur.

3.3 Effect of Water Quality on Water Usage

We demonstrate the influence of cycles of concentration, and thus water quality, on on-site water consumption through an example. Recall that the worse the water quality, the more the sediments in the water and thus the need to blow down (i.e. “flush”) the water pipelines of the cooling system of the data center to avoid damages.

³See Scells et al. [32] for the settings of these experiments. Note, we do not re-run their experiments: we use their values for our water consumption models.

⁴Which is the PUE of our reference data center, and also that used by Scells et al. [32].

Table 1: Power consumption, emissions and water consumption of IR research over the lifetime of a possible experiment across a number of common IR models. Stages in the pipeline (i.e., model training, indexing, and searching) are reported individually; cumulative values across the full pipeline are shown in bold. Running time, power and emission values are taken from Scells et al. [32].

Experiment	Running Time (hours)	Power (kWh)	Emissions (kgCO ₂ e)	Water (L)
BM25 Indexing	0.0809	0.0021	0.0016	0.0108
BM25 Search	0.0025	0.00006	0.00005	0.0003
	0.0834	0.0022	0.0017	0.0113
LambdaMART Training	0.0285	0.0008	0.0006	0.0041
LambdaMART Rerank + BM25	0.0628	0.0017	0.0013	0.0087
	0.0914	0.0024	0.0019	0.0123
DPR Training	16.46	6.74	5.16	34.5910
DPR Indexing	2.42	1.04	0.7958	5.3375
DPR Search	0.4141	0.0002	0.0001	0.0010
	19.3	7.78	5.96	39.9285
monoBERT Training	57.43	57.95	44.38	297.4107
monoBERT Rerank + BM25	23.18	10.8	8.27	55.4277
	80.61	68.75	52.65	352.8384
TILDEv2 Training	15.73	6.91	5.29	35.4635
TILDEv2 Indexing	9.44	4.74	3.63	24.3266
TILDEv2 Rerank + BM25	0.0247	0.0007	0.0005	0.0036
TILDE Expansion	11.89	1.04	0.7958	5.3375
	37.08	12.69	9.72	65.1276
docTquery Expansion	760.48	169.06	129.49	867.6489
	785.68	180.71	138.41	927.4389
uniCOIL Training	17.97	7.24	5.54	37.1571
uniCOIL Indexing	3.66	1.95	1.49	10.0078
uniCOIL Search	0.8966	0.0237	0.0182	0.1216
TILDE Expansion	11.89	1.04	0.7958	5.3375
	34.41	10.25	7.85	52.6050
docTquery Expansion	760.48	169.06	129.49	867.6489
	783.01	178.28	136.54	914.9677

For this example we have considered the TILDEv2 model [43] with docTquery expansion [28]; other models show similar trends, although this was the IR model with the largest water consumption in the results of Table 1. We have computed the on-site water consumption (W_{on}) using the values of cycle of concentrations from the two cooling towers A and B mentioned in Section 3.1, where A was located in a private organization and B in a public hospital ($S_A = 2.25$, $S_B = 1.33$): cooling tower A is more water efficient than B. We keep all other values the same as those used for Table 1. For site A, we obtain $W_{on,A}(TILDEv2) = 602.1609$ L, while for site B, we obtain $W_{on,B}(TILDEv2) = 1261.3524$ L. This example materializes the large impact a lower water quality can have on the

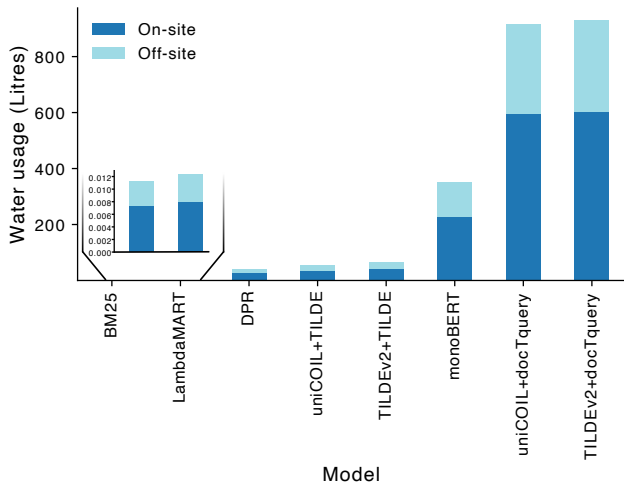


Figure 2: Analysis of on-site and off-site water consumption across different IR models.

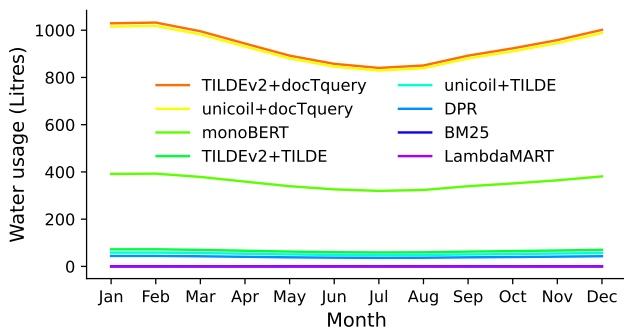


Figure 3: Water consumption of the considered IR models throughout a year for Brisbane, Australia (thus November-February is summer).

amount of water consumed by a data center. The use of high quality water supplies however constitutes an issue per se. First, chemicals could be used to improve water quality: but these add extra costs in the running of the data center operation. These chemicals may also have secondary negative effects on the environment if released into natural water-streams at low down. Chemicals typically used in cooling towers in fact include corrosion and scale inhibitors, algacides and biocides, and pH adjusters, which may have harmful impact on the environment [16] – though we note environmental friendly products do exist, e.g., green corrosion inhibitors [16]. Second, potable water, i.e. drinking water, is typically of high quality – using this water for cooling means subtracting drinking water to the local population, which is problematic in areas with scarce access to sources of potable water.

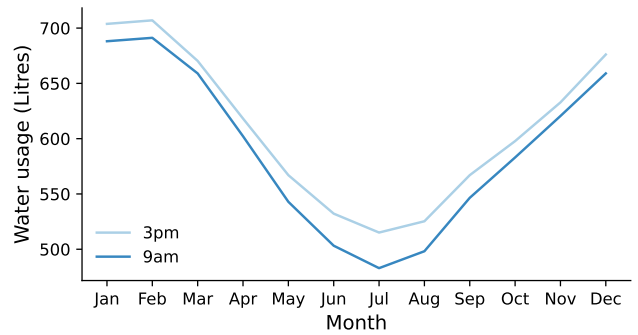


Figure 4: Comparison between on-site water consumption W_{on} for TILDEv2 with docTquery expansion obtained when running the model in the morning (9am) vs. in the afternoon (3pm). The analysis is performed for each month of the year.

3.4 Effect of Time on Water Usage

Next we discuss the effect of time of day and season on water consumption. This effect is due to the rise of the wet-bulb temperature that is typically associated to the warmer part of the day, e.g., afternoons, or to the warmer seasons, e.g., summer. This can be observed clearly in Equation 3: as the temperature $T_w(t)$ raises, so does $WUE_{on}(t)$ and thus consequently W_{on} .

To materialize the effect of temperature changes on (on-site) water consumption, in Figure 3 we report the total water usage W of each IR model throughout the year. We base these results on the same settings used to generate the estimations in Table 1, but instead of taking the annual mean wet-bulb 3pm temperature for Brisbane, we consider the *monthly* mean wet-bulb 3pm temperature. Note that Brisbane is in the Southern Hemisphere, and thus summer is in the period December through to February, and winter is in the period June through August. The figure shows that for the most “thirsty” IR model considered, TILDEv2 with docTquery expansion, water consumption can vary of 192 liters: running this model in the hottest month consumes 23% more water than running it in the coolest month. On the other hand, the impact of season on water consumption for models like BM25 and LambdaMART is minimal (at least in absolute terms).

We perform a similar analysis for showing the impact of time of day. For this, we limit the analysis to the model with highest water consumption, TILDEv2 with docTquery expansion, as an example. We consider two times of the day, 9am and 3pm, and the month with largest difference in mean temperatures at these times in Brisbane: July. The mean 9am wet-bulb temperature in July is 53.6F, and at 3pm is 57.02F: thus the difference is 3.42F. Assuming all other variables are set to the values used for Table 1, the on-site water consumption of TILDEv2 with docTquery expansion at 9am is $W_{on}(TILDEv2, July@9am) = 482.90$, while at 3pm is $W_{on}(TILDEv2, July@3pm) = 515.1$: a difference of 32.2 liters. Figure 4 extends the comparison between the mean 9am and 3pm on-site water consumption for TILDEv2 with docTquery to all months (for each month, we consider the mean over the last 16

years⁵). The figure suggests that for Brisbane the difference in water consumption obtained when running the model at 9am instead than at 3pm is higher during the winter months, than during the summer months. This is because in Brisbane temperature differences between mornings and afternoons are higher in winter than in summer. We note that these findings are location-dependent.

4 DISCUSSION AND CONCLUSION

AI and IR models naturally consume electricity and, as a result, may produce emissions. However, electricity alone is not the only limited resource these models consume. In training or operating IR models, water usage is an important factor to take into consideration, especially when based on increasingly computationally demanding large neural models. Globally, water is increasingly becoming a scarce resource, especially high-quality, potable water⁶.

A reduction in energy or emissions does not necessarily translate into a reduction in water consumption. This is because, in the context of IR models, water consumption does not only depend on the amount of energy consumed (and thus heat generated that needs to be dissipated through cooling) but also on the temperature, humidity and wind conditions of the environment, which naturally fluctuate with times of day and seasons. This aspect is important because it suggests that strategies that reduce CO₂ emissions do not necessarily obtain a reduction of water consumption: indeed, the opposite may occur. This is the case for example for solar power: while this method of electricity production emits no CO₂, it is most efficient in times of the day and seasons with high solar irradiation – which in turn are associated with high temperatures and thus higher quantity of water required to cool the data centers (Section 3.4).

With this paper, we aim to raise awareness about the water consumption associated with large IR models so that practitioners can be conscious of their impact and take steps to minimize water consumption. To this end, we have presented a method for quantifying the water usage of IR methods and compared several models in terms of not only their energy usage and emissions production, but also in terms of their water usage. We have shown that while water consumption of keyword-matching and learning to rank models is minimal, the water consumption associated with neural models can be significant. The analysis we have provided in this paper comes with limitations, first and foremost because power, water and emissions calculations were performed using estimated energy usage and average temperatures.

To further help the IR community to be conscious of the impact of their research on the environment and monitor their energy and water consumption along with their CO₂ emissions, we have developed: (1) a web-based calculator where researchers can conveniently insert the parameters associated to their models and data centers, and compute CO₂ emissions and water consumption, and (2) a plug-in for the Weights & Biases telemetry tool⁷ that allows tracking energy and water consumption, and CO₂ as experiments are run. This material along with other resources associated to the paper are available at <https://github.com/ielab/green-ir>.

⁵Due to the availability of this data from the local meteorology agency.

⁶<https://www.un.org/en/climatechange/science/climate-issues/water>.

⁷<https://wandb.ai/>, an MLOps tool for performance visualization and experimental tracking of machine learning models.

Acknowledgments. This research is funded by the Australian Research Council Discovery Project DP210104043.

REFERENCES

- [1] ISO/TC 207/SC 5. 2006. *ISO 14040:2006 Environmental Management – Life Cycle Assessment – Principles and Framework* (second ed.). ISO, Geneva, Switzerland.
- [2] Susanne Albers. 2010. Energy-Efficient Algorithms. *Commun. ACM* 53, 5 (2010), 86–96.
- [3] Angela H Arthington, Anik Bhaduri, Stuart E Bunn, Sue E Jackson, Rebecca E Tharme, Dave Tickner, Bill Young, Mike Acreman, Natalie Baker, Samantha Capon, et al. 2018. The Brisbane declaration and global action agenda on environmental flows (2018). *Frontiers in Environmental Science* 6 (2018), 45.
- [4] Lotfi Belkhir and Ahmed Elmeligi. 2018. Assessing ICT Global Emissions Footprint: Trends to 2040 & Recommendations. *Journal of Cleaner Production* 177 (March 2018), 448–463.
- [5] Roi Blanco, Matteo Catena, and Nicola Tonello. 2016. Exploiting Green Energy to Reduce the Operational Costs of Multi-Center Web Search Engines. In *Proceedings of the 25th International Conference on World Wide Web*. 1237–1247.
- [6] Fiona Brocklehurst. 2021. *International Review of Energy Efficiency in Data Centres for the Australian Department of Industry, Science, Energy and Resources*. Technical Report. Ballarat Consulting.
- [7] Matteo Catena. 2015. Energy Efficiency in Web Search Engines. In *Sixth BCS-IRSG Symposium on Future Directions in Information Access (FDIA 2015)* 6. 1–2.
- [8] Matteo Catena, Ophir Frieder, and Nicola Tonello. 2018. Efficient Energy Management in Distributed Web Search. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 1555–1558.
- [9] Matteo Catena, Craig Macdonald, and Nicola Tonello. 2015. Load-Sensitive CPU Power Management for Web Search Engines. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 751–754.
- [10] Matteo Catena and Nicola Tonello. 2015. A Study on Query Energy Consumption in Web Search Engines.. In *Proceedings of the 6th Italian Information Retrieval Workshop*.
- [11] Matteo Catena and Nicola Tonello. 2017. Energy-Efficient Query Processing in Web Search Engines. *IEEE Transactions on Knowledge and Data Engineering* 29, 7 (2017), 1412–1425.
- [12] Gobinda Chowdhury. 2012. An Agenda for Green Information Retrieval Research. *Information Processing & Management* 48, 6 (Nov. 2012), 1067–1077.
- [13] Zhuyun Dai and Jamie Callan. 2019. Context-Aware Sentence/Passage Term Importance Estimation for First Stage Retrieval. *arXiv preprint arXiv:1910.10687* (2019). arXiv:1910.10687
- [14] Göran Finnveden, Michael Z. Hauschild, Tomas Ekvall, Jeroen Guinée, Reinout Heijungs, Stefanie Hellweg, Annette Koehler, David Pennington, and Sangwon Suh. 2009. Recent Developments in Life Cycle Assessment. *Journal of environmental management* 91, 1 (2009), 1–21.
- [15] Thibault Formal, Benjamin Piwowarski, and Stéphane Clinchant. 2021. SPLADE: Sparse Lexical and Expansion Model for First Stage Ranking. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Virtual Event Canada, 2288–2292.
- [16] LKMO Goni and Mohammad AJ Mazumder. 2019. Green corrosion inhibitors. *Corrosion inhibitors* 30, 4 (2019).
- [17] Sebastian Hofstätter, Markus Zlabinger, and Allan Hanbury. 2020. Interpretable & Time-Budget-Constrained Contextualization for Re-Ranking. In *24th European Conference on Artificial Intelligence (Frontiers in Artificial Intelligence and Applications, Vol. 325)*. IOS Press, Santiago de Compostela, Spain.
- [18] Leila Karimi, Leeann Yacuel, Joseph Degraft-Johnson, Jamie Ashby, Michael Green, Matt Renner, Aryn Bergman, Robert Norwood, and Kerri L Hickenbottom. 2022. Water-energy tradeoffs in data centers: A case study in hot-arid climates. *Resources, Conservation and Recycling* 181 (2022), 106194.
- [19] Pengfei Li, Jianyi Yang, Mohammad A Islam, and Shaolei Ren. 2023. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. *arXiv preprint arXiv:2304.03271* (2023).
- [20] Jimmy Lin and Xueguang Ma. 2021. A Few Brief Notes on DeepImpact, COIL, and a Conceptual Framework for Information Retrieval Techniques. *arXiv preprint arXiv:2106.14807* (2021). arXiv:2106.14807
- [21] Jimmy Lin, Rodrigo Nogueira, and Andrew Yates. 2021. Pretrained transformers for text ranking: Bert and beyond. *Synthesis Lectures on Human Language Technologies* 14, 4 (2021), 1–325.
- [22] Tie-Yan Liu et al. 2009. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval* 3, 3 (2009), 225–331.
- [23] Sean MacAvaney, Franco Maria Nardini, Raffaele Perego, Nicola Tonello, Nazli Goharian, and Ophir Frieder. 2020. Expansion via Prediction of Importance with Contextualization. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1573–1576.

- [24] Antonio Mallia, Omar Khattab, Torsten Suel, and Nicola Tonellotto. 2021. Learning Passage Impacts for Inverted Indexes. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1723–1727.
- [25] Bhaskar Mitra, Nick Craswell, et al. 2018. An introduction to neural information retrieval. *Foundations and Trends® in Information Retrieval* 13, 1 (2018), 1–126.
- [26] David Mytton. 2021. Data centre water consumption. *npj Clean Water* 4, 1 (2021), 11.
- [27] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. MS MARCO: A Human Generated Machine Reading Comprehension Dataset. In *CoCo@ NIPS*.
- [28] Rodrigo Nogueira and Jimmy Lin. 2019. From Doc2query to docTTTTTquery. (2019), 3.
- [29] World Health Organization et al. 2019. *Progress on household drinking water, sanitation and hygiene 2000-2017: special focus on inequalities*. World Health Organization.
- [30] Yingqi Qu, Yuchen Ding, Jing Liu, Kai Liu, Ruiyang Ren, Wayne Xin Zhao, Daxiang Dong, Hua Wu, and Haifeng Wang. 2021. RocketQA: An Optimized Training Approach to Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 5835–5847.
- [31] Stephen Robertson, Hugo Zaragoza, et al. 2009. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends® in Information Retrieval* 3, 4 (2009), 333–389.
- [32] Harrison Scells, Shengyao Zhuang, and Guido Zuccon. 2022. Reduce, Reuse, Recycle: Green Information Retrieval Research. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2825–2837.
- [33] Jonathan Spinoni, Gustavo Naumann, Hugo Carrao, Paulo Barbosa, and Jürgen Vogt. 2014. World drought frequency, duration, and severity for 1951–2010. *International Journal of Climatology* 34, 8 (2014), 2792–2804.
- [34] Jonathan Spinoni, Jürgen V Vogt, Gustavo Naumann, Paulo Barbosa, and Alessandro Dosio. 2018. Will drought events become more frequent and severe in Europe? *International Journal of Climatology* 38, 4 (2018), 1718–1736.
- [35] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 3645–3650.
- [36] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulkshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, et al. 2022. Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239* (2022).
- [37] Nicola Tonellotto. 2022. Lecture Notes on Neural Information Retrieval. *arXiv preprint arXiv:2207.13443* (2022).
- [38] Shuai Wang, Harrison Scells, Bevan Koopman, and Guido Zuccon. 2023. Can ChatGPT write a good boolean query for systematic review literature search?. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- [39] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. Taxonomy of risks posed by language models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 214–229.
- [40] Lee Xiong, Chenyan Xiong, Ye Li, Kwok-Fung Tang, Jialin Liu, Paul N. Bennett, Junaid Ahmed, and Arnold Overwijk. 2020. Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval. In *International Conference on Learning Representations*.
- [41] Xiang Zhang, Nengcheng Chen, Hao Sheng, Chris Ip, Long Yang, Yiqun Chen, Ziqin Sang, Tsegaye Tadesse, Tania Pei Yee Lim, Abbas Rajabifard, et al. 2019. Urban drought challenge to 2030 sustainable development goals. *Science of the Total Environment* 693 (2019), 133536.
- [42] Shengyao Zhuang, Linjun Shou, and Guido Zuccon. 2023. Augmenting Passage Representations with Query Generation for Enhanced Cross-Lingual Dense Retrieval. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- [43] Shengyao Zhuang and Guido Zuccon. 2021. Fast Passage Re-ranking with Contextualized Exact Term Matching and Efficient Passage Expansion. *arXiv preprint arXiv:2108.08513* (Sept. 2021). [arXiv:2108.08513](https://arxiv.org/abs/2108.08513)
- [44] Shengyao Zhuang and Guido Zuccon. 2021. TILDE: Term Independent Likelihood model for Passage Re-ranking. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 1483–1492.