Lukas Sommer, Nikola Tadic, Raneem Kolakkaden
(Technische Hochschule Nürnberg Georg Simon Ohm)

ColonySearch:
Decentralized Information Retrieval via Ant Colony Optimization

Access to online information is controlled by a handful of corporations whose secret algorithms decide what billions of people see. This problem relates to broader UN Sustainable Development Goals (SDGs: the UN's 17 global targets for sustainable development), especially SDG 9.1 on resilient infrastructure and SDG 16.10 on access to information. One alternative is to spread the search index across many independent computers instead of one central server, so no single company controls what you find. This raises a clear research question: without someone in charge, how does a search query find the right computer without asking every single one?

Earlier systems tried this but failed because they had no smart way to navigate, often as-king every computer just to find relevant results, which gets worse the more computers join. ColonySearch solves this using Ant Colony Optimization (ACO), a method inspired by how ants find food: each connection between computers gets a score that rises when it leads to good results and fades otherwise, so the network learns from experience without any central coordinator. Combined with keyword matching and meaning-based search at each computer, the system finds relevant results on its own. We also developed ACO+MGCHMO (Multi-Goal Cluster Heuristic with Memory), which additionally remembers what topics each computer specializes in to find results even faster.

To establish a controlled baseline, we simulated 20 such computers on a single machine, each holding a slice of 1,574 documents, and tested it with 800 realistic search queries in a 5-fold cross-validation, a standard method to ensure results are not coincidental. Result quality is measured by NDCG@3, a widely used search metric scoring relevance and ran-king of the top three results from 0.0 to 1.0. ACO+MGCHMO achieved 0.729 ± 0.013, nearly identical to asking every computer at once at 0.730 ± 0.014, while contacting only 3.8 computers per search instead of all 20, an 81% reduction in the flooding problem earlier systems faced.

These results confirm technical viability at prototype scale. Robustness against malicious actors and real-world scale remain open questions for future work.