Podcast Name: *ACM ByteCast*
Episode: Ricardo Baeza-Yates

Welcome to the *ACM ByteCast* podcast, a series from the Association for Computing Machinery! The podcast features conversations with researchers, practitioners, and innovators at the intersection of computing research and practice about their experiences, lessons learned, and visions for the future of computing. In this episode, host Juan Miguel de Hoya is joined by Ricardo Baeza-Yates, a leading researcher in information retrieval, search, and Responsible AI, to discuss how modern AI systems depend on the quality and reliability of the information they retrieve. Ricardo's current research focuses on evaluating AI by examining failures, harm, and risk rather than simply measuring success, since errors can have vastly different consequences depending on the context. He argues that language models do not truly understand, reason, or think; they predict likely text, which can produce convincing but incorrect information. This is dangerous in areas such as medicine and law, where users may not have the expertise to recognize errors. Ricardo believes that search and Responsible AI must therefore work together to provide AI agents with accurate, trustworthy information before they generate answers.

The conversation also explores the limitations and social consequences of increasingly replacing traditional search with AI-generated responses. Ricardo warns against excessive "cognitive offloading" onto AI, particularly among younger people who may fail to develop independent writing and thinking skills. Retrieval-augmented generation (RAG) is a promising way to connect AI with information retrieval by giving models relevant source material before generating answers, but Ricardo emphasizes that RAG still cannot prevent all errors because the underlying model remains predictive. Combining reliable search with AI can be highly effective, but relying on AI's predictions alone creates significant risks.

Ricardo's unconventional path into computing began with early education, influential teachers, and his discovery of the mathematical and logical beauty of algorithms. His work on searching large datasets, including the Oxford English Dictionary project, naturally developed into research on information retrieval and web search. He co-authored *Modern Information Retrieval*, which became the field's most cited textbook, eventually leading to leadership roles at Yahoo, Intent, Northeastern University, and U.com.

Ricardo argues that companies deploying AI systems should be held responsible for harmful outcomes, just as manufacturers are responsible when their products fail. AI systems can become biased or dangerous without intentional design because they lack genuine understanding and common sense. He also explains that AI is shifting computer science from solving problems to verifying solutions. A healthier information ecosystem should preserve classical search, fact-checking, and reliable knowledge bases, while developing better ways to verify AI-generated information. In closing, Ricardo advises future AI developers to understand the limitations of data and recognize that AI systems predict patterns rather than truly understand people or reality. He recommends following ACM's responsible computing principles, emphasizing that systems must be ethical, legal, scientifically grounded, developed by technically and domain-qualified experts, and properly authorized. He also argues that Latin America brings valuable perspectives to computing because it exposes the limitations of

wealthy countries' assumptions about "democratizing" AI. Global diversity in languages, cultures, and talent is essential for meaningful innovation, resilience, and progress.

Links:
Learn more about Ricardo Baeza-Yates.
Learn more about the ACM ByteCast podcast at https://learning.acm.org/bytecast.