Roman Kagan

Technical Skills

Expertise in:

- Advanced software engineering principles, including Clean Architecture for system structuring and Clean Code practices for maintainability and readability
- Cloud computing services and architectures, with a focus on efficient change and release management, and implementing CI/CD pipelines for streamlined software deployment
- Programming paradigms such as Functional Programming and Object-Oriented Programming (OOP), as well as building modular systems through Component-Based Software Engineering (CBSE)
- Software development methodologies, particularly Extreme Programming (XP) and Agile frameworks like Scrum and Kanban, alongside virtualization technologies and comprehensive understanding of the Software Development Life Cycle (SDLC)
- Design and development of Software as a Service (SaaS) applications, specializing in Reactive Systems that are fault-tolerant and scalable to ensure high availability and performance

Programming Languages:

- Modern languages: Java, Kotlin, C#, Scala, Python, Rust
- Traditional and systems languages: C, C++, Rust, Bash and ZSH
- Web and data interchange formats: HTML, JSON, XML, FHIR, HL7
- API and Web Service Technologies: RESTful APIs, SOAP, Web Services
- Database querying language: ANSI SQL, OQL

Various Cloud platforms:

- Amazon AWS suite (EC2, EMR, ELB, RDS, Aurora, SQS Messaging, Lambda, Open Search Service, Transcribe, DynamoDB)
- Google Cloud Platform Services (Compute Engine, AppEngine, Kubernetes, IAM, BigQuery, Google Storage)
- Yandex Cloud (API Gateway, Certificate Manager, Cloud Console, Cloud Functions, Compute Cloud, Container, Data Streams Service for managing data streams in real time, DataLens, Managed Service for Apache Airflow, Managed Service for Apache Kafka, Managed Service for ClickHouse, Managed Service for Elasticsearch)
- Hetzner baremetal
- Logging and data visualization: Kibana, LogStash, Fluentd, Splunk
- Streaming and messaging systems: Kafka, Spark
- Distributed systems coordination: Hadoop, Zookeeper
- Container and Virtualization services: Docker, Kubernetes, Terraform, VMWare NSX

Information Retrieval Systems and Database Management Systems (DBMS):

- Search Platforms: Solr, LucidWorks, ElasticSearch, Vector Search, OpenSearch, Coveo, Lucene, Yandex (45GB source code reverse engineered)
- Databases: Oracle, SQL Server, PostgreSQL, MySQL, Percona, MariaDB, CockroachDB, RockDB, DataStax Cassandra, ClickHouse, InfluxData I0x
- Search Algorithms: BM25, TF/IDF, PageRank, LTR (Naive Bayes Classifier, RankNet, RankBoost, CatBoost, XGBoost)
- Search Observability and Metrics: NDCG, MRR, MAP, Kibana, Graphana, Prometheus, OpenTelemetry, OpenCensus, OpenTracing, CNCF
- Prompt Engineering Tools: OpenAI's ChatGPT 4, Perplexity, Copy.ai, HuggingFaces, Mistral

Multiple Development Frameworks and Libraries:

- Micronaut and Spring Framework modules
- Data processing frameworks: Apache Spark and Apache Flink
- API documentation: Swagger, Confluence, GraphQL Apollo, Jira
- Build and project management tools: Gradle, Maven, Ant with Ivy
- Java EE/Jakarta EE components: JDBC, JMS, JMX, JPA, Hibernate

Development and Collaboration Tools:

- Integrated Development Environments (IDEs): IntelliJ IDEA, PyCharm, RustRover, XCode, Jupyter, Visual Studio

- Version control and CI/CD: Git, GitLab, GitHub Actions, Jenkins
- Workflow automation and monitoring: Airflow, Ansible Tower, AppDynamics

Professional Experience

McKesson   2019 – Present
Senior Software Engineer – AI & Prompt Engineering

- Contributed to a major initiative aimed at enhancing the coding capabilities of McKesson's healthcare data management systems, focusing on improving the accuracy and utility of healthcare data through rigorous code and data reviews.
- Actively reviewed and audited healthcare datasets, identifying and rectifying errors, bugs, and inconsistencies. This process significantly enhanced the reliability of data used across various McKesson healthcare applications.
- Worked closely with data scientists, clinical staff, and IT teams to develop and implement new guidelines and frameworks for data handling and processing. This collaboration ensured that all data enhancements met the stringent requirements of healthcare standards.
- Led efforts to develop and refine datasets specifically tailored for healthcare use cases, involving complex data structures and requiring a deep understanding of clinical environments.

Kroger   2018 – 2019
Senior Software Engineer – Search & Ranking

- Utilized AWS S3 and AWS EMR for analyzing clickstream and add to cart signals and metadata catalog.
- Developed AWS SNS producers that read data from external RESTFul web services, transformed it, and sent it to topics (written in Java with Kotlin)
- Led the development of an analytics engine to aggregate click engagement and other signals for training a machine learning framework to improve data quality
- Developed dashboards for monitoring search relevance for onboarding new items and out of stock items, allowing merchandisers to adjust orders from suppliers and taxonomists to adjust item classifications
- By reviewing and suggesting enhancements to search ranking, Kroger saw a $500 million boost in annual sales and witnessed an overall improvement in customer experience and satisfaction.

Albertson Companies   2017 – 2018
Senior Software Engineer – Search & Ranking

- Spearheaded the enhancement of query expansion capabilities, employing AI algorithms to enrich search queries with synonyms, related terms, and predictive user behavior analytics. This pivotal upgrade led to a 45% improvement in search accuracy and a 30% increase in user engagement.
- Architected and deployed an innovative knowledge graph system. This system leveraged structured data and intricate entity relationships to deliver a 40% boost in search result relevance and contextuality, greatly enhancing user experience.
- Initiated and guided the development of a predictive search feature that intelligently suggested search terms and anticipated user queries based on popular trends and individual search history. This feature significantly reduced search time and improved user satisfaction.
- Masterminded the integration of Solr, Cassandra, and Spark to create a state-of-the-art analytics engine. Utilizing Jupyter Notebooks and Python, this project allowed for in-depth data analysis, which directly contributed to a 50% reduction in decision-making time for search optimization strategies.
- Played a key role in designing and developing a high-performance, scalable web application using ASP.Net with C#, setting a new benchmark for application efficiency within the company.
- Directed the ambitious redevelopment of a major segment of the existing platform from JSPs to Spring Boot with Kotlin, which included establishing a robust CI/CD pipeline on Pivotal Cloud Formation. This initiative resulted in a 25% faster deployment cycle and a 20% increase in application performance.
- Successfully integrated Spring Boot web services with Okta's JWT security framework. This strategic move fortified the application's security infrastructure, mitigating risks and vulnerabilities by 30%.

- Innovated new ETL jobs in Scala for Spark, enabling efficient data loading from Kafka topics to DSE Cassandra with Solr. This breakthrough enhanced data processing speeds by up to 40%, significantly optimizing search engine performance.
- Offered expert consultancy on data modeling and schema best practices for search engines, alongside recommendations for enhanced monitoring and alerting mechanisms, reinforcing the foundation for superior search engine reliability and efficiency.

Walmart Labs   2015 – 2017
Senior Software Engineer – Search Relevance and Observability

- Wrote BM25 implementation as a Solr plugin, at that time Solr 4.x only had TF/IDF implementation and Walmart needed a better ranking algorithm and we knew from publications that OKAPI BM25f is the way to go
- Developed a recommendation system to find the best matches of VUDU streaming videos, integrated with Walmart's search platform, using Java, interfaced with Solr, ElasticSearch, and BigQuery in GCP
- Created and optimized a pipeline that processes billions of telemetric events (a la Google Analytics) in real time using Kafka, Spark, and Cassandra
- Developed a sophisticated analytics engine for detecting and fixing data anomalies using Linear Regression and ARIMA statistical models
- Utilized BigQuery, Hive, and Hadoop Data Lakes to analyze customer behavior, demographics, preferences, catalogs, categories, segments, and product attributes to optimize assortment for each store, department, and customer.
- Wrote a highly performant Normalized Discounted Cumulative Gain (NDCG) algorithm to use in A/B tests, measure relevance changes, enhancing predictive models for ontology, synonyms, and whitelists/blacklists
- Worked on search personalization where search results used based on the user's past behavior, preferences, and context where search history, clicks, and interactions.
- Worked on semantic search to understand the intent and context of the search query. That involves using natural language processing and semantic analysis to interpret the meaning of the queries to return more relevant results.
- Incorporated user feedback loop, such as ratings, reviews, click data, and add to cart signals to refine and improve search algorithms and relevance.
- Created Faceted Search to allow users to narrow down search results by applying facets based on specific attributes and categories.
- Optimized Content to ensure that content is well-structured, tagged, and optimized for search engines to improve its visibility and relevance in search results
- Fixed a NullPointerException issue in the QueryResultKey of Solr on the Lucene layer, and submitted the fix to the Apache Committee for review (https://tinyurl.com/ybdwvta2)

For the sake of brevity, earlier experiences that have been omitted can be located on LinkedIn

Education

- Bachelor of Science in Computer Science, Minsk Polytechnic University
- Continuing Education in Computer Science, University of Michigan, Engineering Department
- "Search Fundamentals" by CoRise (Grant Ingersoll and Daniel Tunkelang)
- "Machine Learning Power Search" By Doug Turnbull (getsphere.com)
- "Solr - Think Like A Relevance Engineer" - Open Source Connections
- Participated in "Information Retrieval WS 22/23, Lecture 01" (YouTube Channel)