International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Scholarly Publisher
RS Global Sp. z O.O.
ISNI: 0000 0004 8495 2390

Dolna 17, Warsaw,
Poland 00-773
+48 226 0 227 03
editorial_office@rsglobal.pl

ARTICLE TITLE	AN ARTICLE ON DATAMINING IN GAMES AND THE SCHEMA SYSTEM
DOI	https://doi.org/10.31435/ijitss.4(48).2025.4474
RECEIVED	23 September 2025
ACCEPTED	19 November 2025
PUBLISHED	26 November 2025
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

AN ARTICLE ON DATAMINING IN GAMES AND THE SCHEMA SYSTEM

Duger Gantuya *(Corresponding Author, Email: d.gantuya@mnun.edu.mn)*
Associate Professor, Head of the Department of Economics, Mongolian National University

Tsyrenzhapov Pavel Sergeevich
Undergraduate, Mongolian National University

ABSTRACT

Data mining, the process of discovering patterns and knowledge from large amounts of data, has become an indispensable discipline in the 21st century. While its conceptual roots and the term itself emerged in the late 1980s and 1990s, the exponential growth in data generation and computational power has propelled it to the forefront of technology. Data mining is not a single method but a collection of techniques and algorithms used to transform raw data into valuable information and actionable insights. This practice has found a critical application in the video game industry, giving rise to the field of "game analytics." The ability to collect data on player behavior on a massive scale, especially with the rise of online and multiplayer games, allows developers to understand their audience, balance gameplay, and improve the user experience iteratively. This article provides an overview of data mining, traces its historical progression, and discusses its importance in modern game development. It concludes with a concrete example: an analysis of the schema system in Valve's Source 2 engine, which exemplifies how a modern game engine is architecturally designed to facilitate data-driven development and analysis.

KEYWORDS

Data Mining, Game Analytics, Player Behavior, Schema System, Source 2 Engine, Modular Design

CITATION

Duger Gantuya, Tsyrenzhapov Pavel Sergeevich. (2025). An Article on Datamining in Games and the Schema System. *International Journal of Innovative Technologies in Social Science*, 4(48). doi: 10.31435/ijitss.4(48).2025.4474

In this 21st century, the amount of demand for data usage is increasing rapidly at a very exponential pace. Along with it, the processing capabilities of computers and increase in growth of internet-capable devices with each day and week has made it easy to gather even more and more data about every aspect of our daily lives. But, this data needs to made useful, morphed into information and knowledge to be of real value and of use for making decisions. This requires adequate techniques and tools that help uncover hidden and valuable information within the vast space of patterns in the data. Otherwise, it has no deeper meaning. That's what data mining is for, to get the data and then to give it form, knowledge and reason to help us understand it from the source. While data mining has always been there, the attention increase and the term itself have been only recent things, appearing in the late 80s-early 90s. Data mining is not a single technique, but a chaos assembly of methods, techniques, algorithms, research processes, think pieces applied in an iterative or explorative process.

21st century is not only considered to be the age of data mining itself; it has also found it's way into game production and has become a crucial part of game development and maintenance. This has led to creation of a new field named "game analytics" – application of analytics to game research & development. It uses data mining techniques to discover various patterns and extract information from game-related data, like player behavior data. Emergence of game datamining and analytics may be attributed to a coincidence of several developments.

This growing interest in data mining and analytics has been made possible by several key developments and advances. The widespread adoption of Internet-capable gaming devices allows developers nowadays to track player behavior remotely and without any big intrusion risks on a large scale. Before that, play testing

has had to be done in-house while developers observed play testers in a laboratory-like setting while they are playing the game. This was at a much smaller scale and did not allow to test for large-scale predictions, as a small group of testers is not representative of a large playing base. Today, as gaming becomes a mainstream phenomenon, played by an increasingly diverse audience, games need to be made to appeal to a wide array of people. In this case, data mining is a valuable tool for gathering data at this big of a scale. This heavily applies to very large types of games like massively multiplayer online games or social network games. They are played by hundreds and thousands of players simultaneously, allowing for various degrees of online interactions. Remote data collection makes gathering of such large-scale and long-term datasets even possible, even more so with rapid increase of game development budgets. Integration of data science into game development is a continuous process that in the end becomes a great tool that can provide developers real-time feedback and allow for iterative improvements over the lifespan of a game.

One of these developers who makes data mining in their recent titles possible, allows for iterative development, live debugging and ability to implement changes as fast as possible with data-driven techniques is Valve. The game engine that powers their recent games is Source 2, and it has been built to be a modular engine, continuing the tradition established with the original Source Engine. Modular game engine design out of things like modules, components, etc. allows the engine developers, for example, to swap out implementations of a module for another, or provide multiple components, add new ones and make custom ones. This is the perfect example of modularity in programming that is very crucial for engine development. To help facilitate data mining, development and analysis, majority of engine's modules provide serialized reflection data built from C++ classes with compilers, and it's name is the Schema system. Schema system creates very detailed descriptions of classes, enumerators and types. A huge majority of classes have a schema binding with detailed descriptions of their members and inheritance. The schema system is powered by a module with the same name, and it's the core of this whole system that maps out the data and describes almost everything about the engine. It has various functions like: find or create type scopes, create scopes for modules, find declared classes and so on. Classes described by the Schema system also describe very classes that it uses to describe other classes. In game development, schema data is used to provide serialized reflection data to debugging tools live, that can visualize everything printed out from the engine output in all detail in a debugger like a physics visualizer to test physics of the world, manipulate it and debug it in the end.

Schema system is also used to power the data-driven editors and tools to create new resources that are used by the engine for various things like material, model, and generic data editors. Schemas provide all the class information that editors use to display so they could function, provide metadata such as property descriptions, what kind of resources a parameter is allowed to use, exposed script bindings for tools like a visual node editor, what it can do, while in return, the editor can access exposed schema enumerators.

The community, on the other hand has developed tools that can extract and create schema headers for all the exposed classes and enumerators residing within a game that can be used within tools that have direct access to the game such as a cheat or a mod, or it can be used as a data mining tool to verify what was changed between the updates to the engine/game, check for difference between the produced results and achieve various conclusions from it. In the end, the schema system ends up becoming useful for both developers and the community outside, allowing for easier game development with data/schema-driven document-producing tools, visualization of game/engine data from the schemas live in the visual snoopers/debuggers, and by the community outside to: create new stuff that has access to the game at runtime, simplified process of making a mod for a game and in data mining purposes: document all of the new changes that may have occurred in the schemas between each major game update, check for critical differences and create and visualize data to: track progress of future game installments, analyze the produced data in tools like websites that would track data of each played match, and so on.

REFERENCES

1. https://www.games.cg.jku.at/wp-content/uploads/2022/10/chapter1_GDABook.pdf, A Brief Overview of Data Mining and Analytics in Games
2. https://ioaglobal.org/blog/role-of-data-science-in-gaming-industry/, The Role of Data Science in the Gaming Industry
3. Game Data Mining: Clustering and Visualization of Online Game Data in Cyber-Physical Worlds, sciencedirect.com
4. https://www.irma-international.org/viewtitle/77753/?isxn=9781466643017 Modular Game Engine Design
5. https://praydog.com/reverse-engineering/2015/06/24/source2.html An Analysis of the Source 2 Engine: The Schema System