

Seznam Diskuze: Jak funguje tým, který se stará o miliony komentářů?

17.5.2024 - Eva Juraszková | Seznam.cz

Baví vás diskutovat s lidmi v online prostoru? Pokud ano, možná vás bude zajímat, kdo se stará o to, aby diskuze na seznáckých a dalších webech správně fungovaly a dokázaly si poradit s miliony příspěvků.

Proto jsme se Jiřího Kupky, který šéfuje vývoji v projektu Seznam Diskuze, zeptali, jak vypadá práce jeho týmu a co ho na ní nejvíc baví.

Představ nám Diskuze, o co se jedná?

Projekt Seznam Diskuze vznikl před necelými čtyřmi lety jako podpůrná služba pro naše obsahovky. Chtěli jsme dát uživatelům prostor popovídat si o aktuálních tématech přímo na službě a tak vznikl diskuzní embed – jednoduchá komponenta, kterou je možné vložit kamkoliv do stránky. Diskuze jsme nejdřív nasadili na web Proženy, následovala Garáž, potom naše velké zpravodajské weby – Novinky a Seznam Zprávy. Momentálně embed najdete na deseti našich obsahovkách a asi stovce webů externích partnerů, což s sebou přináší výrazné množství trafficu. V databázi máme asi 70 milionů komentářů, skoro miliardu lajků a naše API musí zvládnout několik tisíc requestů za sekundu.

S jakými technologiemi pracujete?

Na backendu používáme Python 3, pro komunikaci s databází pak SQLAlchemy, data vydáváme frontendu pomocí GraphQL API, kde nám pomáhá například framework Strawberry. Frontend píšeme v JavaScriptu. Používáme Node.js, React nebo seznamácký framework IMA.js. Data máme uložena primárně v MySQL, sekundárně pak pro rychlý přístup k datům používáme Redis. Všechny aplikace nám běží v prostředí Kubernetesu, o který se stará náš SCIF (Seznam Compute Infrastructure) tým.

Tým Seznam Diskuzí se v práci rozhodně nenudí.

Jaká část práce tě baví nejvíc a proč?

Na práci mě nejvíc baví nejrůznější výzvy, kterých na tomhle projektu řešíme velké množství. Z pohledu uživatele se sice jedná o jednoduchý embed, který vykresluje kartičky komentářů, ale na pozadí se skrývá množství procesů a technologická komplexita. Uživatelé denně napíší asi 70 000 komentářů a přidají zhruba 700 000 lajků. Během exponovaných událostí je to ještě víc. Každý jeden komentář u nás prochází kontrolou, jestli neobsahuje výzvy k násilí nebo jiným způsobem neporušuje podmínky používání služby. Tuto kontrolu primárně provádí AI a příspěvky, se kterými si nedokáže poradit, propadají k našim editorům, kteří je posuzují. Naše administrace jim například nabízí nástroje, pomocí kterých nedochází k duplicitní práci, když je editorů na směně více.

Také se rádi pouštíte do nových výzev? Staňte se posilou Jirkova týmu.

Využíváte AI i v jiných oblastech?

Jak už jsem nakouzl, strojové učení je další zajímavá oblast, se kterou se u nás člověk setká. Pracujeme na několika klasifikátorech textu, které nám posuzují nezávadnost příspěvku nebo zkoumají, zda se v textu nachází nějaká dezinformace a jaká. Pracujeme taky na doporučování příspěvků – snažíme se například komentáře v diskuzích seřadit od těch nejkvalitnějších, protože na našich obsahových službách chceme mít primárně kultivovanou diskuzi. Vzhledem k trafficu, který neustále roste, musíme myslet na to, ať jsou dotazy na API optimální a stejně tak databázové schéma. Změny se při tomhle množství dat dělají dost špatně.

Co vás čeká v nejbližších měsících?

Diskuze se staly nedílnou součástí našich obsahových služeb a v průběhu posledních čtyř let jsme v Seznamu rozvíjeli sociální prvky a možnosti interakce s uživatelem. Dva roky jsme například experimentovali s veřejným profilem uživatele, prošlapali jsme si nějaké slepé, ale i nadějně cestičky a v příštím roce se chystáme spustit novou generaci služby Lidé.cz. Na „novém Liděčku“ pak uživatelé najdou svůj profil, budou moci psát příspěvky podobně jako na Twitteru nebo zakládat skupiny jako na Facebooku. Chystáme i pár zajímavých a nevšedních produktových fičur, takže se uživatelé mají na co těšit.

Jak funguje tvůj tým i po lidské stránce?

Náš tým se dnes skládá z 11 lidí. Část týmu je v Ostravě, část v Praze. Navzdory vzdálenosti nám to spolu pěkně klape. Minimálně jednou za kvartál se snažíme všichni potkat – někdy v Praze, jindy v Ostravě. Vidíme se nejen v kanclu, ale po práci občas hrajeme deskovky, jdeme si zahrát badminton nebo prostě jdeme na pivo.

Jsou nějaké plány na rozšíření týmu?

Ano. Ve stávajícím množství lidí a komponent se stane, že nám občas uteče nějaká chyba nebo skřípe nějaký proces, který by se dal urychlit nebo zrefaktorovat. Hledáme proto QA engineera/DevOpsáka, který by komplexně dbal na kvalitu našich aplikací a procesů. Kolegu, který má rád, když všechno šlape jako hodinky. Máme například sdílené knihovny, u kterých je potřeba, nejlépe automatizovaně, kontrolovat, že jsou ve všech projektech v aktuálních verzích. Samostatnou kapitolou jsou pak testy. Ty si u nás píše každý vývojář sám, ale je potřeba připravit kvalitní testovací framework a prostředí pro psaní testů. Důležitý je pro nás taky CI/CD proces, který by měl být co nejrychlejší, protože u případného bugfixu se každá minuta počítá. Víme, že kvalitu aplikace je důležité zajišťovat i u uživatele, a to především pomocí metrik nebo logů. I tady v posledních měsících vidíme velkou potřebu neustále ladit naše Grafana a Kibana boardy. Takže posilu rozhodně uvítáme.

<https://blog.seznam.cz/2024/05/seznam-diskuze-jak-funguje-tym-ktery-se-stara-o-miliony-komentaru>