

# Recommender Systems @ Seznam.cz

**Ondřej Javornický**  
Product manager

**Matěj Jakimov**  
Researcher/ML Engineer



# What?

- Who?
- Product side
- Engineering and research side
- Questions (after each section)



# Why product and engineering side?

- RecSys
  - context is important
  - implementation and infrastructure is important
- bad infrastructure → expensive/impossible experiments and analysis
- good engineering → good research
- engineering + research → ability to **do things right**
- good product management → ability to **do right things**



# Who?

- Ondřej Javornický
  - In Seznam.cz from 2011
  - 2013 - 2016 responsible for advertising system and ad targeting
  - In 2014 he set up the Recommender Systems Department
- Matěj Jakimov
  - Completed this course at 2014
  - Research and engineering on recommender systems at GaussAlgo → Seznam



**Product side**



Stream.cz

Zavřít ×

**Nová epizoda seriálu Přijela pout: Pavlinka**

Intelektuální slušná dívka na návštěvě u světských. Během jediného dne se stane svědkem nevěry, zhárství a vyhrožování se zbraní. A tím její soužení zdaleka nekončí!

20:03

Seznam Zprávy · Pondělí 29. října. Svátek má Silvie.

**Šťastné pondělí: Prosba čínskému vedení: Nechovte se zase k Zemanovi tak, jak se chová on k nám**

Přežili jsme 28. říjen. Premiér Babiš se potkával s evropskými státníky, prezident Zeman odměnil své...

Zákulisí ceremonie na Hradě: David upustil medaili, Kubera se nedostal do sálu  
Merkelová oznámila konec v čele CDU. Kanceltřkou zůstane maximálně už jen tři roky

Novinky

**Meteorologové rozšířili výstrahu před silným větrem**

Varování Českého hydrometeorologického ústavu (ČHMÚ) před silným větrem platí nově na celém území republiky...

Nové hlavní nádraží se má jmenovat Šalingrad, rozhodli Brňané  
V Ralsku našli mrtvou včičku. Zabilo ji zřejmě divoké prase  
Milované i nenáviděné boty zvané kačenky jsou opět v módě  
Osm vět, které muže děsí dřívě, než je žena vysloví  
Jen 99 kusů a ani o jeden víc. Alpina představila parádně upravené BMW

Super

**Ukázala odhalené tělo bez podprsenky a dočkala se hejtů: Fandové byli na Terezu Kerndlovou hodně sprostí**

Chtěla potěšit svoje fanoušky, ale dočkala se pravého opaku. Zpěvačka Tereza Kerndlová na Instagram...

Chtěla by maminku, která vypadá jako princezna? Mahulena Bočanová dceři přivedla jako náhradnici Kateřinu Brožovou

Ve 14 letech se dostala do semifinále modelingové soutěže! Jedna z nejkrásnějších hereček

Email Email Profi

Založit nový e-mail

Zobrazit přihlášení do Emailu

Služby

|               |             |                  |              |
|---------------|-------------|------------------|--------------|
| Akce / Letáky | Dovolená    | Lidé             | Prohlížeč    |
| Auto / Moto   | Email Profi | Mapy             | Reality      |
| Bazar         | Hry         | Mobilní aplikace | Volná místa  |
| Deníky        | Jízdní řády | Pohádky          | Zboží / Móda |

Počasí · Česká republika

Předpověď větru

|  |              |  |              |  |              |
|--|--------------|--|--------------|--|--------------|
|  | <b>12 °C</b> |  | <b>13 °C</b> |  | <b>12 °C</b> |
|  | Odpoledne    |  | Večer        |  | V noci       |

Reklama · Koupit reklamu

ASTRATEx

BESTSELLERY

SLEVA AŽ

35%

PROHLÉDNOUT



Firmy

Přidat firmu

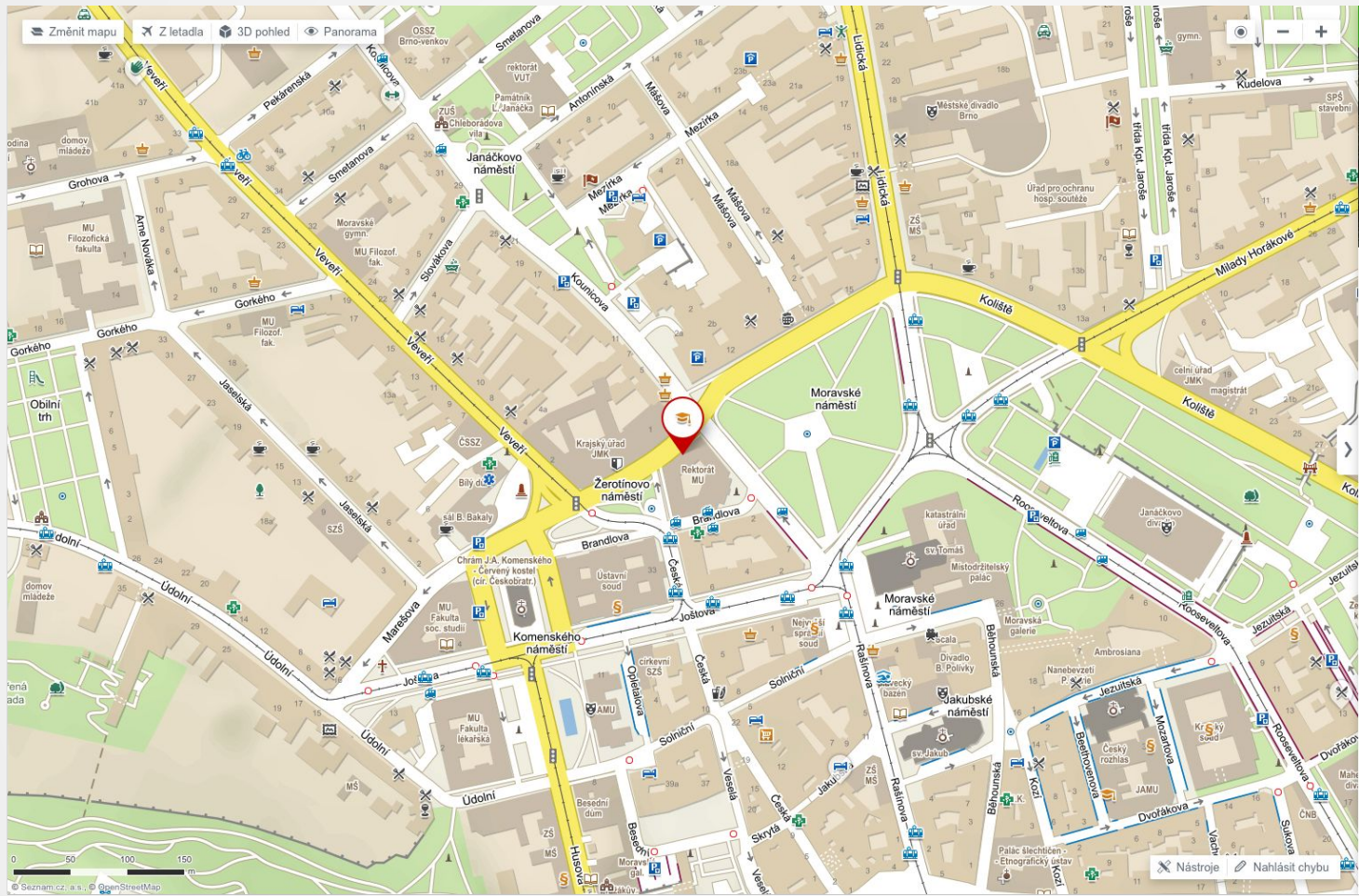
**Mountfield**

Prodej hobby potřeb

Web · Trasa · 571 613 343

Zašovská 132, Valašské Meziříčí

|            |             |          |              |
|------------|-------------|----------|--------------|
| Autobazary | Kadernictví | Penziony | Restaurace   |
| Dovolená   | Kosmetika   | Pizzerie | Rozvoz lístů |



PANORAMA FOTKY 3D POHLED

### Masarykova univerzita

Vysoká škola, univerzita

Žerotínovo náměstí 617/9  
602 00 Brno-město  
Jihomoravský kraj

- Do plánování
- Uložit
- Sdílet
- Tipy na výlet

E-mail: [info@muni.cz](mailto:info@muni.cz)  
 Telefon: +420 549 491 111  
 Web: [www.muni.cz](http://www.muni.cz)  
 Hodnocení: ★★★★★ 1 recenze - hodnotit

Univerzita nabízí studium na devíti fakultách humanitního, přírodovědeckého a sportovního zaměření. Zúčastňujeme se soutěží o podporu výzkumných projektů. Klademe důraz na mezinárodní spolupráci se zahraničními univerzitami a vědeckými institucemi. Zajišťujeme příležitosti v přístupu ke vzdělání.

Partneři:

Další odkazy:  
[www.firmy.cz](http://www.firmy.cz)



## Zprávy dne



FOTO: Profimedia.cz

### Meteorologové rozšířili výstrahu před silným větrem

**29. 10. Aktualizováno** Varování Českého hydrometeorologického ústavu (ČHMÚ) před silným větrem platí nově na celém území republiky od pondělních 17 hodin. Na hřebenech hor může vítr v nárazech dosahovat rychlosti až 110 kilometrů v hodině. [Celý článek »](#)



### Policejní prezident Tuhý končí ve funkci, má nahradit Klausova v Bratislavě

**29. 10.** Spekulace o tom, že se působení Tomáše Tuhého (46) ve funkci policejního prezidenta po čtyřech a půl letech chýlí ke konci, jsou nyní...



### Nové hlavní nádraží se má jmenovat Šalíngrad, rozhodli Brňané

**29. 10.** Brno - Šalíngrad. Právě tak by se podle Brňanů, kteří se zapojili do ankety, mělo jmenovat nové železniční nádraží. Fádní pojmenování...



### V Ralsku našli mrtvou vlčici. Zabilo ji zřejmě divoké prase ▶ Video

**29. 10.** Mrtvé tělo mladé vlčice našli lesníci v lesích bývalého vojenského újezdu Ralsko na Českolipsku, nedaleko obce Hradčany. Podle ochránců...

REKLAMA

REKLAMA

**MITSUBISHI**

**SUV EVOLUCE**



**LEGENDÁRNÍ**

**OUTLANDER**

**jž od 608 450 Kč**



**MITSUBISHI**  
**MOTORS**

## 🕒 Stalo se

11:50



### Muž podezřelý z vraždy prodavačky na Třebíčsku se sám přihlásil policii

Čtyřicetiletý muž, který v sobotu v Moravských Budějovicích na Třebíčsku podle policie ubodal svou bývalou přítelkyni, se na třech...

REKLAMA



**ALTISPORT**

**ZIMNÍ BUNDY**



# ONE MAN SHOW

Tohle je ta největší SHOW na českém internetu!  
Nic podobného tady nikdy nebylo.  
Internetový diktátor Kazma Kazmitch  
společně se svou bandou zlosynů dobývá svět.

celkem zhlédnutí  
78,135,897

odběratelů  
65,100

★ Odebírat

## Epizody

Vše

2. generace

1. generace

Bonusy

### 2. generace



#### 29. Odhalení zákulisí natáčení ONE MAN SHOW 1/10

472,245 zhlédnutí středa 19. září



#### 26. Nejdiskutovanější kauza Kazma vs. Mareš bude mít nečekané odhalení

403,356 zhlédnutí pondělí 10. září



#### 23. Nový díl je na spadnutí a překvapí celou Českou republiku

687,598 zhlédnutí 10. května 2017



#### 20. Kazma je zpátky a bude to zase velký!

362,006 zhlédnutí 7. března 2017



#### 17. Dacjukova brusle – Teaser

502,942 zhlédnutí 29. čerence 2016



#### 28. Song Cizí zeď od Kazmovy ONEMANSHOW Foundation

629,061 zhlédnutí čtvrtek 13. září



#### 25. Prostřeno! Making of: Zákulisí největšího televizního skandálu v Česku

1,199,352 zhlédnutí 27. května 2017



#### 22. Takto v Praze probíhalo natáčení největšího podvodu na Justina Biebera

681,783 zhlédnutí 22. března 2017



#### 19. Dacjukova brusle – Making of

438,348 zhlédnutí 27. čerence 2016



#### 16. MS v hokeji 2016 – Making of

909,703 zhlédnutí 22. čerence 2016



#### 27. Odhalení skandálu prohrané sázky o Ferrari za 8 000 000 Kč

1,982,933 zhlédnutí středa 12. září



#### 24. Prostřeno! Odhalení největšího televizního skandálu v Česku

4,367,320 zhlédnutí 16. května 2017



#### 21. Nový díl plný neuvěřitelných zvrátů skončil žalobou od Justina Biebera

2,334,105 zhlédnutí 15. března 2017



#### 18. Dacjukova brusle

1,414,436 zhlédnutí 19. čerence 2016



#### 15. MS v hokeji 2016

2,487,606 zhlédnutí 8. čerence 2016

MS 2018  
 NHL  
 Euro Hockey Tour

Tipsport extraliga  
 Kontinentální liga  
 Ostatní

Reprezentace  
 První liga

## Sparta - Brno 4:2

Sparta slavila před nejvyšší návštěvou sezóny výhru nad Brnem

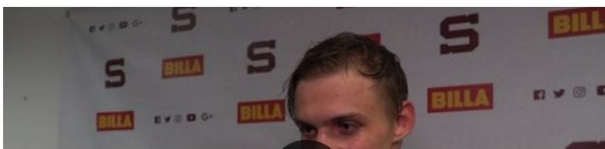
| Minuta | Tým | Hráč    | Akce |
|--------|-----|---------|------|
| 3.     |     | Köhler  | 0:1  |
| 9.     |     | Mueller | 0:2  |
| 10.    |     | Delisle | 1:2  |
| 34.    |     | Smejkal | 2:2  |
| 37.    |     | Blain   | 3:2  |
| 38.    |     | Kumstát | 4:2  |



[Záznam online reportáže »](#)

21. října 2018 17:54 - Praha

Hokejisté Sparty díky dokonalému obratu zdolali ve 12. kole extraligy mistrovskou Kometu Brno 4:2 a připsali si čtvrtou výhru v řadě. Hosté naopak poprvé v sezóně vyšli ve dvou duelech za sebou bodově naprázdno. Teprve druhé vítězství Sparty z posledních devíti vzájemných duelů s Kometou sledovala v O2 areně nejvyšší návštěva této sezóny extraligy 15 471 diváků.



EXKLUZIVNĚ NA DIGI TV **DIGI**

# Nejlepší světový fotbal

Premier League  
 BUNDESLIGA  
 SERIE A  
 LaLiga

[Chci objednat DIGI TV >](#)

Reklama

EMAIL

HRY.cz

 PROŽENY

FIRMY.CZ

Kupi.cz

SUPER.CZ

SBAZAR.CZ



GARÁŽ.CZ



Televize Seznam

S BLOG

HOROSKOPY.CZ

LIDÉ

SPORT.CZ

MAPY.CZ

Novinky.cz

S REALITY.CZ

Zboží.cz

stream.cz

S KLIK.cz

S AUTO.CZ

POČASÍ.CZ

S | ZPRÁVY



# About Seznam.cz

## The most visited web on Czech internet

- 35 millions PV / day
- 3,2 millions RU / day

In November 2017, 6.2 million real users visited Seznam.cz. That same month, the entire Czech internet population was 7.22 RUs.

## Eminent Publisher

Daily publish over

- 300 own articles / videos
- 400 external articles

The screenshot shows the Seznam.cz homepage with a search bar at the top. The main content area features several news articles with images and headlines. On the right side, there is a weather widget showing current conditions and a TV program schedule. At the bottom, there are more news snippets and a copyright notice.



#### Maminka



**Maminka sdílela trik, jak vstávání. Ne každému s Jedním z častých důvodů, fakt, že mu vypadá dudlík...**

#### deníkPlus



**Indie plánuje riskantní misi na jižní pól Měsíce. Proč je nebezpečná?**  
Indická kosmická agentura plánuje již v druhé polovině letošního roku podniknout riskantní vysazení své...

#### Sport



**Moreva pi'o či Supričaba. Helmu Pančochové zlobí originální náplý**

S helmou od svého hlavního sponzora na olympiádě závodit nemůže, a tak aspoň Šárka Pančochová močila...

Ukurná kosal vyvěšel se v Koreji český olympionik. S manželkou se zasnili o druhém Naganu  
Smrt na přejezdu. Chěl si zpešit trénink, zaplatil životem

#### Proženy



**Manželství pro osudu: Které celebrity spolu vydržely jen pár hodin nebo dnů?**  
Zahrádkáři, už je čas! Co si právě teď předpřítovat a je Vyzhiete si obilň? Jednoduché cviky, které vám zvep

#### Novinky



**zve Na Strahově se šikují taxikáři, chystají protestní jízdu Prahou**

Taxikáři se ve čtvrtek sešli na pražském Strahově, odkud hodlají před poledním vyjet do centra metropole na...

Další vzpoura v ČSSD. Vedení strany se postavili odborníci  
Meteorologové varují před náledím

Zeměň novinář Karel Slezák, který v Zemanově štábu napadl reportéry  
Hyundai Santa Fe odhaleno, chce zaujmout obří maskou a úzkými světlomety  
Co ničí důvěru v jinak fungujícím státu

#### Stream



**Moz, který několikrát zrcimoval oslou Evropu, je stále nebezpečný**  
Takže lidi zajímá...



**Nevětší hamba našeho sportu. Takto komunisti 'odř' Slav**



**Opřiní jiní | pohodlí a luxus na prvním**

#### Seznam Zprávy - Čtvrtek 8. února, Svátek má Mláda.



**To nechcete zažít. Dvě a půl hodiny na úřadě a dozvíte se: „Nevim. Neznám. Vyjádřím se do 30 dnů“**

Nabízíme vám bonusový materiál k nové epizodě Zábavní vyšetřování. Jde o prvky záznam z jednání...

Snih už se v Česku neudrží, teploty zůstanou přes den nad bodem mrazu  
At Okamura ve Sněmovně čelí přímé demokracii. Co řekl o tábore v Letech, je sprostáma, miní šel Pirátů

#### Kafe



**10 nejvíc, ne...**

Jan Werich patří k našim neznámějším a nejváženějším divadelním a filmovým legendám, i přesto, že díky polito...

#### Autožvč



**Těchhle pět automobilových klasik nám bude chybět**

První Škoda Octavia RS pomalu ale jistě mizí z českých silnic.

#### Prima Ženy



**5 důvodů, proč se zamilovat do francouzku**

Poibky, zvlášť ty vášnivé a mlínečné, jsou přímou rozbíškou k lásce, a to nejen fyzické, ale i dušev...



**Seznam Zprvy** Čtvrtek 8. února, Světlet mládeže

**To nechte zažít. Dvě a půl hodiny na úřadu a dozvíte se „Nevim, Neznám, Vyjádřím se do 30 dnů“**  
 Nablízem vám hovorový materiál k nové epizodě Záhledného vyšetřování. Jde o první záznam z jednání...  
 Sblíž se s Českou neudí. Nechť objasní přes den nad bodem mrazu...  
 At Okamura ve Sněmovně opět přemlouvá. Co řekl o tábore v Letech, je správně...  
 min 547 Práto

**Noviny**  
**Na Strahově se šikají taxikáři, chystají protestní jízdu Prahou**  
 Taxikáři se ve dnech sedli na problem Strahov, odčůd hodjí jízdu poodem vjezd do centra metropole na...  
 Další vopura v ČSD. Vedení strany se postaví odbojní...  
 Bude povně ukaže emonace kříd, v nebyly vyvolaly odstavce motry...  
 Zemřel novinář Karel Slezák, který v Zemeřel Šarha nebyl nebyly...  
 Myndra Šarha Fe odhaleno, chce zaognout drfí mapou a užijmí odstavce...  
 Nový jíz podotok: osazenka mže vyvolat jen neznámí

Ukurní kosal vjezdí se v Koreji Český olympionk. S manželk Smrt na přejezdu. ČHM si zpsařf trénnk, zapletí životem

**Super**  
**Jana Prošková odvolává významné kulturní: Přivkáží a g rasiloval manžel a proč odmlná tlevení roků**  
 Nablízem mřívava hračka a hřídla trní. Jak vyhrnout vjezdí křídla a jak osadit letadlo do zrakobory...  
 Úbez záměru je tu rozurný křídla. Provozka Praveka Šarova se postává potoků letů avně...  
 Nejhrdnější porokně SuperStar: Už jen kvidi letě nádrně Slovence bude stát za to se...  
 na nové letě avně

**Stream**  
 Mzr, který nřikolník ztecimoval ostou Evropu, je stále nebezpečný...  
 Největší hraha našeho sportu. Taklo komunisti...  
 Úplní jiní pohodi a luxus na poním...  
 Seznam Zprvy Čtvrtek 8. února, Světlet mládeže

**Novinky**  
**Protest Taxikářů hodjí p**  
 Další vopura v ČSD. Vedení strany se Meteorologové vranjí před nádom...  
 Zemřel novinář Karel Slezák, který v Zemřel Šarha Fe odhaleno, chce zaognit...  
 Co nří dříve v jnak fungujm vztahu

**TV program**  
**Převě 6h**  
 10:30 11:00 Omluva: Ivan Bartoš  
 10:00 10:45 F. L. Věk (913)  
 10:00 11:05 Soudobím kámošf: Třapsko  
 10:25 10:55 Fénka pro hojy ČT sport

**Walmartská inspirace**  
 Človk Kean Stern...  
 Človk Chase W EDP 75...  
 Človk Wellington...  
 Človk Carver...  
 Človk 02-414...  
 Človk 0-3 280 Kč

**Proby**  
**Podívte česká citulka: Lahodná křemová pokrývka vte došane**  
 Citulka je jeden z hlavních přířf České kuchyně, pteho je...  
 serno a sobe nedostaněte... vřídno u šarame je jaco...  
 Měšabf pro ostou: Které celebrity jsou vydržely jen pár hodin nebo dnů?  
 Záhledný avně: Úž je čas! Co si právě teď předpřetovat a jak...  
 Vyžehete si oboj? Jednoduché cviky, které vám zpev...

**Auto žvř**  
**Těchhle pět bude chybět**  
 První Škoda Octavia...  
**První ženy**  
 5 důvodů, proč...  
**Libat „po francouzku“**  
 Polibky, zvlášť v vášnivě a mlenecké, jsou přímou...  
 rozbuškou k lásce, a to nejen fyzické, ale i duševně...

**Auto žvř**  
**VW dává šedesát avněm avčuaněmu logu, už přefi Golf no nedostane**  
 Už 16 let neprodlo logo Volkswagenu blabou emonku, a to...  
 je doopravdy i na prvky avně komerčních avně...

**Maminka**  
**Maminka sdělila trik, jak uklidit v noci dítě bez vřívání. Na každému se to šik**  
 Jevník v Českých avně, před se dvě v noci poodit, je...  
 fakt, že mu vypadne duška. Zazmvo vřídno rodní poodit...

**Krpké haly**  
**Oldřich Kaiser se vřídil ve velkém stylu. Nová partnerka mu pomohla, dokonce s nři zpřva**

**Maminka**



Maminka sdělila trik, jak uklidit v noci dítě bez vřívání. Na každému se to šik...  
 Jevník v Českých avně, před se dvě v noci poodit, je...  
 fakt, že mu vypadne duška.

**denikPlus**



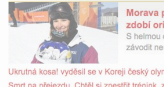
Indie plánuje...  
 Indická komedie...  
 letovního roku

**Prožny**



Ukurní kosal vjezdí se v Koreji Český olympionk. S manželk Smrt na přejezdu. ČHM si zpsařf trénnk, zapletí životem

**Sport**



Morava přifřf Supričab zlobí originální náplvy...  
 S helmou od svého hlavního...  
 zřivodf nemže, a tak asořf!

**Stream**



Mzr, který nřikolník ztecimoval ostou Evropu, je stále nebezpečný...  
 Největší hraha našeho sportu. Taklo komunisti...  
 Úplní jiní pohodi a luxus na poním...

**Novinky**



Protest Taxikářů hodjí p...  
 Další vopura v ČSD. Vedení strany se Meteorologové vranjí před nádom...  
 Zemřel novinář Karel Slezák, který v Zemřel Šarha Fe odhaleno, chce zaognit...  
 Co nří dříve v jnak fungujm vztahu

**Kafe**



10 nejvříc, ne...  
 Jan Wierov patřf k našim nejznámějším a nejváženějším...  
 divadlém a filmový legendám, i přesto, že díky politik...

**Auto žvř**



Těchhle pět bude chybět...  
 První Škoda Octavia...

**První ženy**



5 důvodů, proč...  
 „Libat „po francouzku“...  
 Polibky, zvlášť v vášnivě a mlenecké, jsou přímou...  
 rozbuškou k lásce, a to nejen fyzické, ale i duševně...

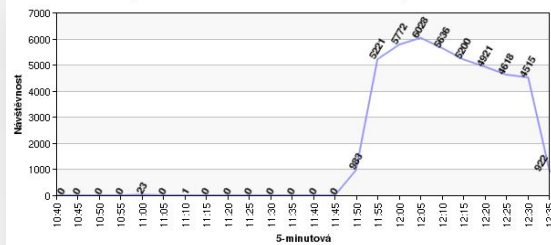


# Pre-recommender era

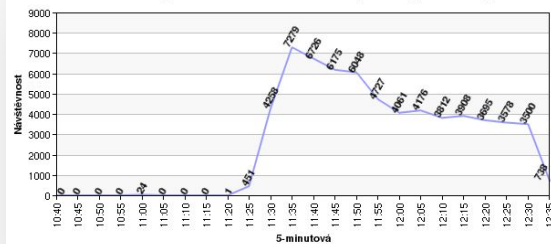
Based on real-time statistic of each article and manual positioning articles on homepage.

- teams of editors
- real-time statistic
- operation 24/7
- manually selecting each article
- manually positioning each article

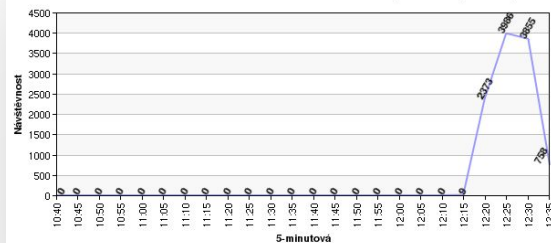
1 - Babiš přinesl na Hrad demisi a odnesl si pověření k sesta...



2 - Za billboardy se Zemanovou knihou padne pokuta, byla to k...



3 - Očima Saši Mitrofanova: Duel: detaily, zásahy, fauly



#### Maminka



**Maminka sdílela trik, jak uklidnit v vstávání.** Ne každému se to líbí. Jedním z častých důvodů, proč se dítě fák, že mu vypadne dudlík. Zatímco větš

#### deníkPlus



**Indie plánuje riskantní misi na jižní pól Měsíce. Proč je nebezpečná?**  
Indická kosmická agentura plánuje již v druhé polovině letošního roku podniknout riskantní výsazní svě...

#### Proženy



**Poctivá česká cibulačka polévka vás dostane**  
Cibule je jeden z hlavních pi sama o sobě nedoobná...

**Manželství pro osudu: Které celebrity spolu vydržely i v Zahradkách, už je častí** Co si právě teď předpřetovat a j Vyžehíte si obilí? Jednoduché cviky, které vám zpev



**Morava pi'o či Suprleza. Helmu Pančochové zdobí originální náplvy**  
Či helmou od svého hlavního sponzora na olympiádě zvodit nemůže, a tak aspoň Šárka Pančochová mohla...

**Ukrutná kosa! Vyřídil se v Koreji český olympionik. S manželkou se zasnil o druhém Naganu**  
Smrt na přejezdu. Chtěl si zpešit trenink, zaplatit životem

#### Novinky



**Na Strahově se štkují taxikáři, chystají protestní jízdu Prahou**

Taxikáři se ve čtvrtek sešli na pražském Strahově, odkud hodlají před polednem vyjet do centra metropole na...

**Další vzpoura v ČSSD. Vedení strany se postavili odborník**

Meteorologové varují před náledím

Zemřel novinář Karel Slezák, který v Zemanově stábu napadl reportéry

Hyundai Santa Fe odhaleno, chce zaujmout obli maskou a úzkými světlomety

Co nidi důvěru v jinak fungujícím vztahu

#### Stream



**Moz, který několikrát zdecimoval ostou Evropu, je stále nebezpečný**  
Takové lidi zajímá...

**Nevětší hanba našeho sportu. Takto komunisti 'odř' Slav**

**Úplní jiní suru, tozy je pohodlí a luxus na prvním**

#### Seznam Zprávy - Čtvrtek 8. února. Svátek má Mláda.



**To nechte zařít. Dvě a půl hodiny na úřadě a dozvíte se: „Nevim. Neznám. Vyjádřím se do 30 dnů“**

Nabízíme vám bonusový materiál k nové epizodě Zvláštní vylétnutí. Jde o prvový záznam z jednání...

Snih už se v Česku neudrží, teploty zůstanou přes den nad bodem mrazu

**At Okamura ve Sněmovně čelí přímé demokracii. Co heki o táboře v Letech, je sprostáma, miní šáf Pirátů**

#### Kafe



**10 nejvíc, ne**

Jan Wierich patří k našim neznámějším a nejváženějším divadelním a filmovým legendám, i přesto, že díky politc...

#### Autožvč



**Těchhle pět automobilových klasik nám bude chybět**

První Škoda Octavia RS pomalu ale jistě mizí z českých silnic.

#### Prima Ženy



**5 důvodů, proč se zamilovat do francouzku**

Polibky, zvlášť ty vášnivé a mlénečné, jsou přímou rozbuzkou k lásce, a to nejen fyzické, ale i dušev...





Maminka



**Maminka sdílela trik, jak uklidnit v noci dítě bez vstávání. Ne každému se to líbí**  
Jedním z častých důvodů, proč se dítě v noci probudí, je fakt, že mu vypadne dudlík. Zatímco většina rodičů prostě...

Proženy



**Poctivá česká cibulačka: Lahodná krémová polévka vás dostane**

Cibule je jeden z hlavních pilířů české kuchyně, přesto je sama o sobě nedooceníná – většinou ji bereme jen jako...

**Manželství pro osudu: Které celebrity spolu vydržely i v Zahradkách, už je časté** Co si právě teď předpřetovat a j Vyzhiete si obilňoj? Jednoduché ovky, které vám zpev

Novinky



**Na Strahově se šikují taxikáři, chystají protestní jízdu Prahou**

Taxikáři se ve čtvrtek sešli na pražském Strahově, odkud hodlají před polednem vyjet do centra metropole na...

**Další vzpoura v ČSSD. Vedení strany se postavili odborníci**

Meteorologové varují před náledím

Zemřel novinář Karel Šlezák, který v Zemanově štábu napadl reportéry

Hyundai Santa Fe odhaleno, chce zaujmout obli maskou a úzkými světlomety

Co ničí důvěru v jinak fungujícím vztahu

Seznam Zprávy - Čtvrtek 8. února. Světek má Mláde.



**To nechcete zažít. Dvě a půl hodiny na úradě a dozvíte se: „Nevim. Neznám. Vyjádřím se do 30 dnů“**

Nabízíme vám bonusový materiál k nové epizodě Zvláštního vylétnutí. Jde o nový záznam z jednání...

Snih už se v Česku neudrží, teploty zůstanou přes den nad bodem mrazu

**AT Okamura ve Sněmovně čelí přímé demokracii. Co heki o táboře v Letech, je sprostáma, miní šaf Pirátů**

Autožvč



**Těchhle pět automobilových klasik nám bude chybět**

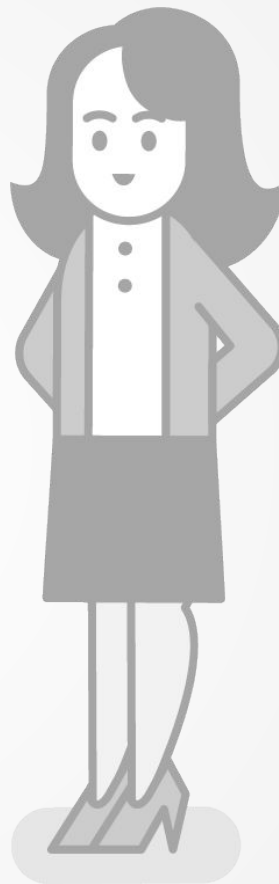
První Škoda Octavia RS pomalu ale jistě mizí z českých silnic.

Prima Ženy



**5 důvodů, proč se francouzské ženy považují za líbat „po francouzsku“**

Polibky, zvlášť ty vášnivé a milenecké, jsou přímou rozbuškou k lásce, a to nejen fyzické, ale i dušev...



deníkPlus



### Indie plánuje riskantní misi na jižní pól Měsíce. Proč je nebezpečná?

Indická kosmická agentura plánuje již v druhé polovině letošního roku podniknout riskantní vysazení své...

Proženy



### Poctivá česká cibulačka: Lahodná krémová polévka vás dostane

Cibule je jeden z hlavních pilířů české kuchyně, přesto je sama o sobě nedooceníná – většinou ji bereme jen jako...

Manželství pro osudu: Které celebrity spolu vydržely jen pár hodin nebo dnů?

Zahradkáři, už je čas! Co si právě teď předpřetovat a jaké triky vám pomůžou?

Vyžehlete si obilěj! Jednoduché cviky, které vám zpevní rýsy a vygumují křupy pod očima

Stream



Moz, který několikrát ztecimoval oslou Evropu, je stále nebezpečný... Takže lidi zajímá...



Největší hanka našeho sportu. Takto komunisti... ,odř Siva



Úplně jiný SUV, tady je pohodlí a luxus na prvním

Seznam Zprávy - Čtvrtek 8. února. Světek má Mláda.



To nechcete zažít. Dvě a půl hodiny na úřadě a dozvíte se: „Nevim. Neznám. Vyjádřím se do 30 dnů“

Nabízíme vám bonusový materiál k nové epizodě Zvláštní vyláštvení. Jde o nový záznam z jednání...

Sníh už se v Česku neudrží, teploty zůstanou přes den nad bodem mrazu

At Okamura ve Sněmovně čelí přímé demokracii. Co heki a táboře v Letech, je sprostáma, mání šáf Pirátů

Kafe



10 nejvíc, ne

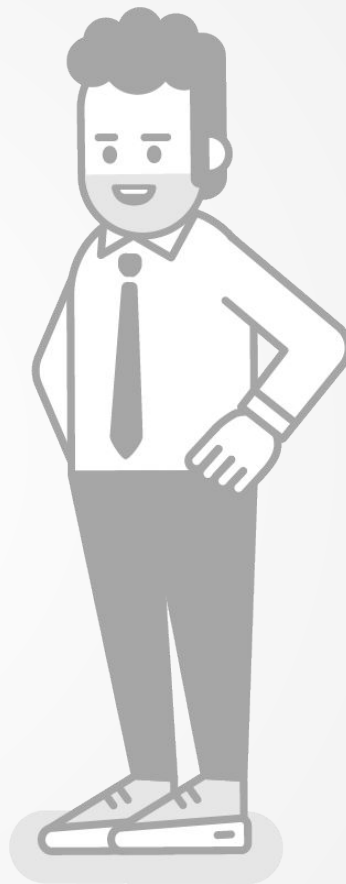
Jan Werich patří k našim neznámějším a nejváženějším divadelním a filmovým legendám, i přesto, že díky polito...

Autožvč



### Těchhle pět automobilových klasik nám bude chybět

První Škoda Octavia RS pomalu ale jistě mizí z českých silnic.



**Proženy**



**Poctivá česká cibulačk polévka vás dostane**  
Cibule je jeden z hlavních pi sama o sobě nedobrodina...

Manželství pro osudu: Které celebrity spolu vydržely i Zahradkáři, už je čes! Co si právě teď předpřestovat a j Vyžehíte si obilěj! Jednoduché ovky, které vám zpev

**Sport**



**Morava pi'o či Suprlezba. Helmu Pančochové zdobi originální náplvy**  
S helmou od světlo hlavního sponzora na olympiádě závodit nemůže, a tak aspoň Šárka Pančochová mohla...

Ukrutná kosa! vyřítel se v Koreji český olympionik. S manželkou se zasnili o druhém Naganu Smrt na přejezdu. Chtěli si zpešití trenink, zaplatit životem

**Novinky**



**Na Strahově se štkují taxikáři, chystají protestní jízdu Prahou**  
Taxikáři se ve čtvrtek sešli na pražském Strahově, odkud hodlají před polednem vyjet do centra metropole na...

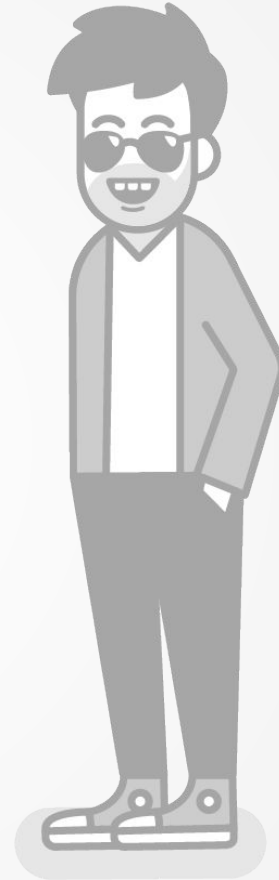
Další vzpouza v ČSSD. Vedení strany se postavili odborníci  
Meteorologové varují před náledím  
Zemřel novinář Karel Slezák, který v Zemanově štábu napadl reportéry  
Hyundai Santa Fe odhaleno, chce zaujmout obli maskou a úzkými světlomety  
Co ničl důvěru v jinak fungujícím vztahu

**Seznam Zprávy** - Čtvrtek 8. února. Světek má Mláde.



**To nechte zažít. Dvě a půl hodiny na úřadě a dozvíte se: „Nevim. Neznám. Vyjádřím se do 30 dnů“**  
Nabízíme vám bonusový materiál k nové epizodě Zvláštní vyšetřování. Jde o nový záznam z jednání...

Snih už se v Česku neudrží, teploty zůstanou přes den nad bodem mrazu  
Ať Okamura ve Sněmovně čelí přímé demokracii. Co heki o táboře v Letech, je sprostáma, miní šáf Pirátů



editor's choice

1<sup>st</sup> recommender



# Our start with recommendation

## Performance

- global popularity

## Personalization

- hiding of the read content
- field of interest - based on advertisements targeting

editor's choice

1<sup>st</sup> recommender



# Our start with recommendation

## Performance

- global popularity

## Personalization

- hiding of the read content
- field of interest - based on advertisements targeting



editor's choice

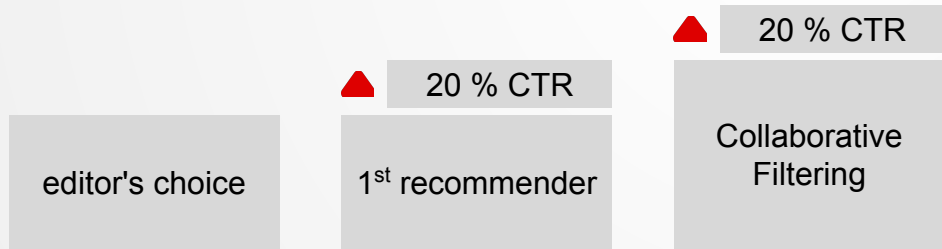


20 % CTR

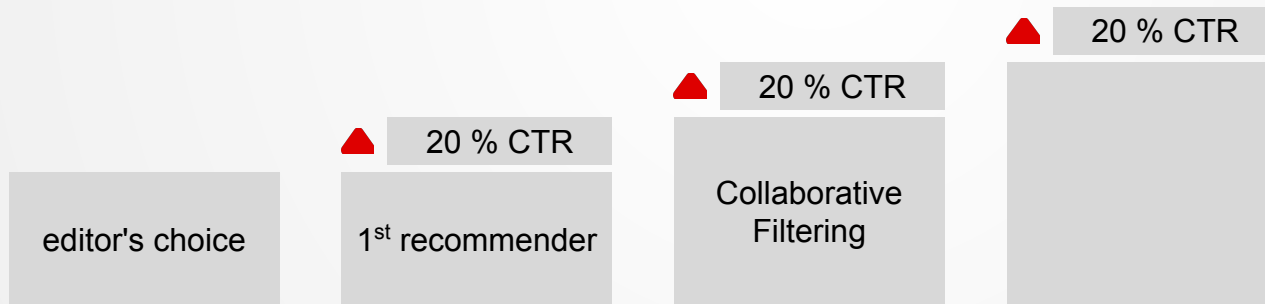
1<sup>st</sup> recommender

Collaborative  
Filtering









#### ✖ Reflex



#### **Kolik bere Moravec? Česká televize to odmítá sdělit. Káže transparentní vodu, pije ale víno**

Dle zprávy z LN 6. 2. 2018 nechce veřejnoprávní Česká televize sdělit veřejnosti platy svých moderátorů. Ar...

#### 🏆 Sport



#### **Trenér Jandač se netěší do olympijské vesnice. Důvod je prostý. Spolubydliči**

Už i druhá část hokejové reprezentace odletěla do dějiště olympijských her. Hráči z extraligy a zbytek realizační...

#### 🌸 Proženy



#### **Dole bez! Proč by ženy během zimy měly doma chodit „naostro“?**

Co kromě huňaté šály, horkého čaje, bavlněných ponožek a zvýšeného příjmu vitamínu C právě teď v zimě potřebujete...

[More articles](#)

## ✖ Reflex



### **Kolik bere Moravec? Česká televize to odmítá sdělit. Káže transparentní vodu, pije ale víno**

Dle zprávy z LN 6. 2. 2018 nechce veřejnoprávní Česká televize sdělit veřejnosti platy svých moderátorů. Ar...

## 🏆 Sport



### **Trenér Jandač se netěší do olympijské vesnice. Důvod je prostý. Spolubydlíci**

Už i druhá část hokejové reprezentace odletěla do dějiště olympijských her. Hráči z extraligy a zbytek realizační...

## 👉 Proženy



### **Dole bez! Proč by ženy během zimy měly doma chodit „naostro“?**

Co kromě huňaté šály, horkého čaje, bavlněných ponožek a zvýšeného příjmu vitamínu C právě teď v zimě potřebujete...

## 👉 Prima Ženy



### **5 důvodů, proč se manželské páry přestávají líbat „po francouzsku“**

Polibky, zvlášť ty vášnivé a milenecké, jsou přímou rozbuškou k lásce, a to nejen fyzické, ale i duševní...

## 📄 Security Magazin



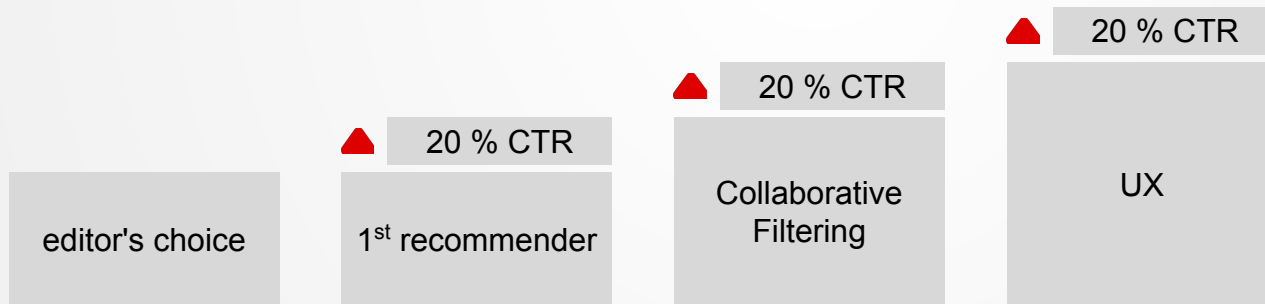
### **Vrah Schimak: Zabil tolik lidí, že už je nikdo nespočítá. Udala ho hostinská, když se probouzel**

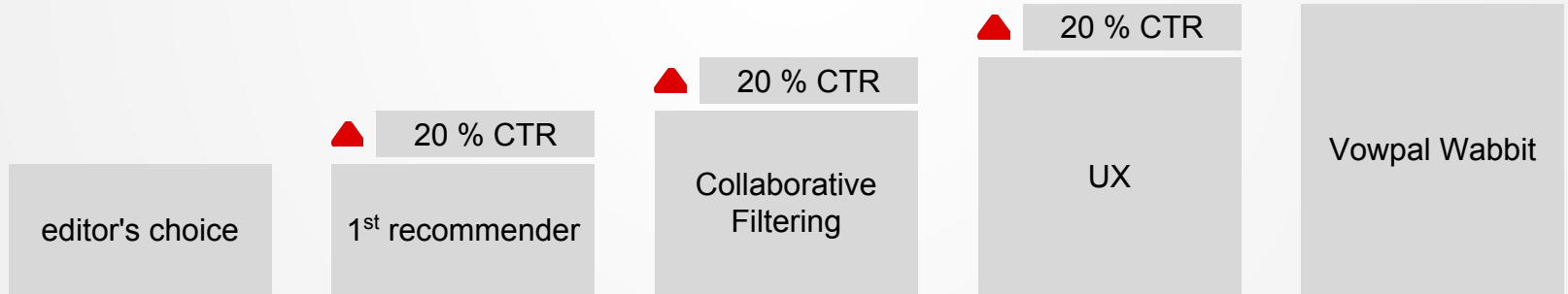
Rádění Antona Schimaka trvalo prakticky pouhý měsíc a půl - od května do června roku 1887. Začalo přepadení...

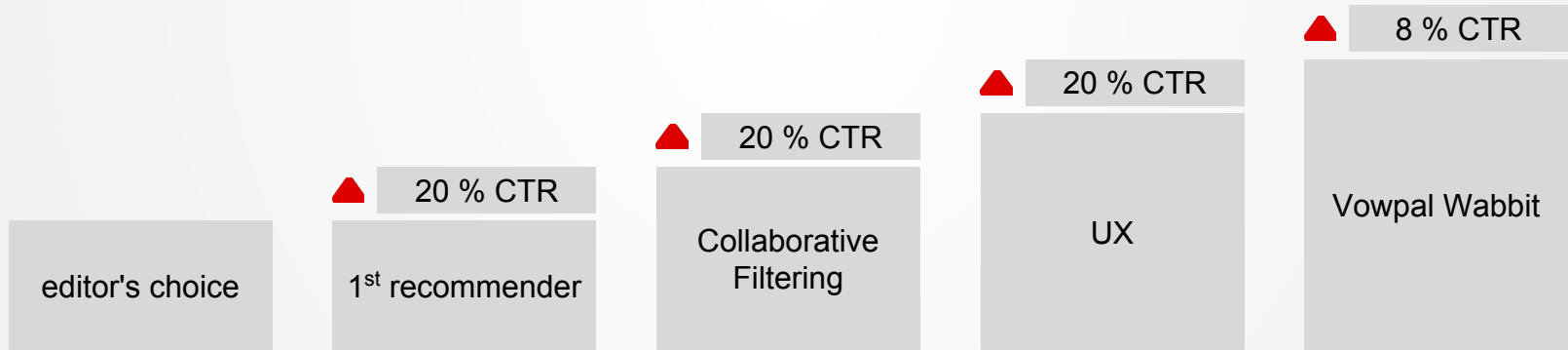
## 📄 denikPlus



### **Indie plánuje riskantní misi na jižní pól Měsíce. Proč je nebezpečná?**



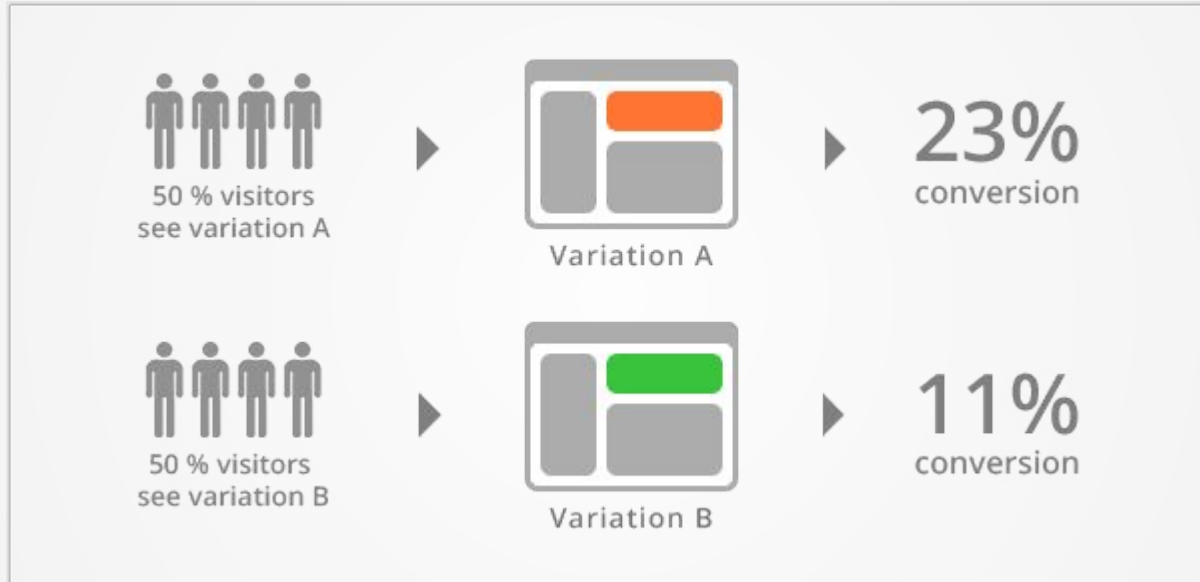




# Evaluation



# Evaluation





# Evaluation

AB testing



# Evaluation

AB testing

- no offline evaluation will tell you what one AB test can



# Evaluation

## AB testing

- no offline evaluation will tell you what one AB test can
- you have to know what exactly you want to test and how you can measure it



# Evaluation

## AB testing

- no offline evaluation will tell you what one AB test can
- you have to know what exactly you want to test and how you can measure it
- at Seznam we run about 10 AB tests on RecSys in same moment and 2-3 test of UI



# Evaluation

User experience



# Evaluation

## User experience

- ask users
  - use forms and user testing



# Current goals

- cookie → registered user
  - for user identification
  - GDPR
- more interactive elements in UI
  - like / dislike, comment, following...
- new metrics
  - oriented on user satisfaction



# Time for questions





# Engineering and research

# Technical point of view

- Primary problem is not an algorithm, but:
  - System must respond
  - ~1000 req/sec/box
  - ~50ms avg latency (100ms max latency)
  - All components - tens of physical machines - High Availability
  - Data are noisy and big



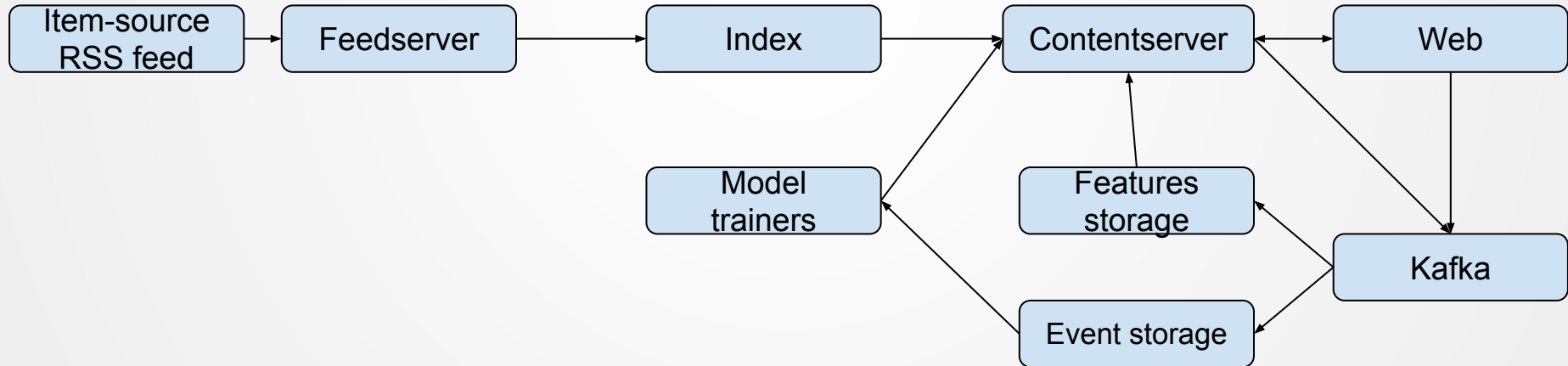
# Our stack

- Python, Docker, Kubernetes, Git, CI/CD, Spark, Scala
  - Ability to quickly develop something is the most important
- MariaDB Galler, Couchbase, Kafka, Elasticsearch, Memcache, Swift Object Store
- Monitoring and alerting
  - Prometheus, Grafana, Influxdb
- → Lot of engineering :) 8 developers + 2 researchers + admins + many other teams
- Scrum ... Constantly under construction



# Architecture

- very simplified, ignoring some important subsystems
- forked for each content service (e.g. Novinky, Stream)
- no precomputed recommendations

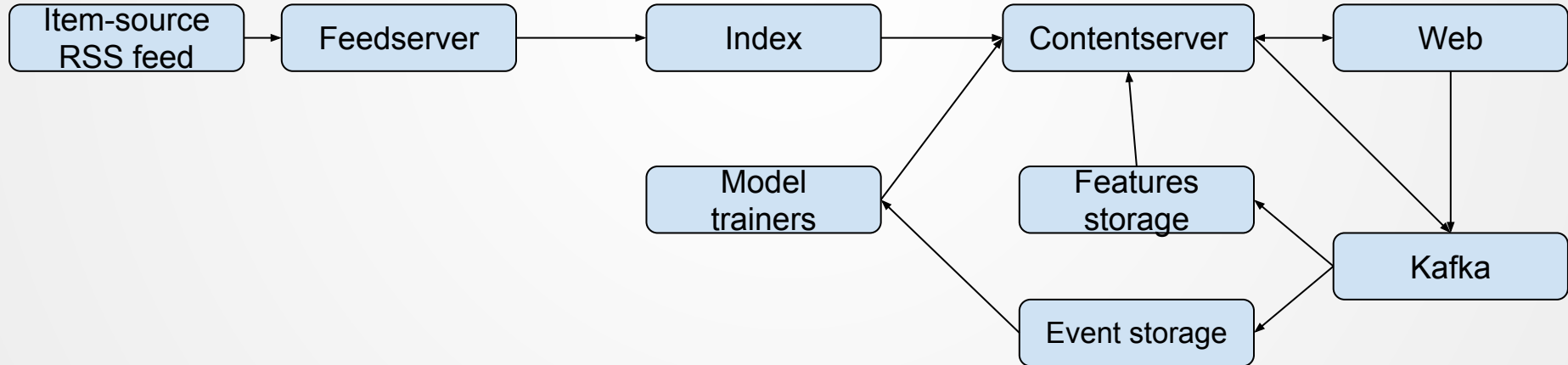


# Time for questions



# Architecture

- (almost) everything quickly reconfigurable (AB testable)
- Controlserver - key ingredient



# Data collection

- Clean data, get to know your data
  - unittests, integration tests
  - post deployment analysis - do not trust (your) code
  - beware of JavaScripts
  - beware of bots (good, harmless, evil)
  - garbage in-garbage out, misleading measurements



# Deployed algorithms

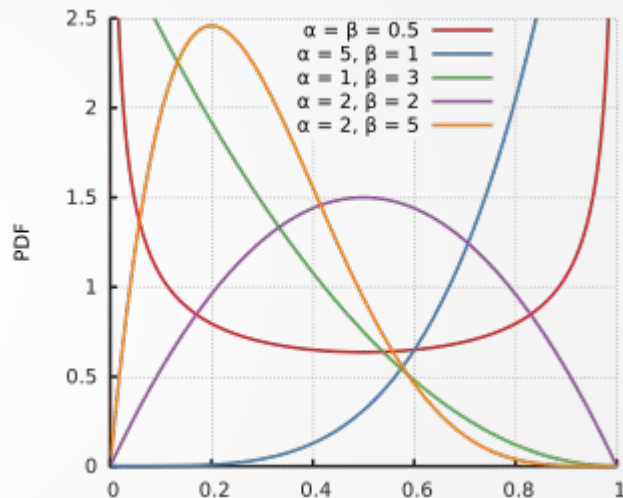
- Matrix Factorization
  - Alternating Least Squares
    - faster than SGD
    - parallelize better
  - Each 10-50 minutes - one run of ALS from scratch -> Couchbase
  - Now - single machine computation
    - 10 min instead of 50 min
    - ~30-100GB RAM, 32 CPU
    - online, more flexible for experiments, need some additional infrastructure





# Deployed algorithms

- Not every user have vector
- New items do not have vector
- Cold start
  - Beta distribution - CTR
  - Thompson sampling
  - Explore-exploit dilemma
  - Clustering on CF vectors (demo)
- Cold start - logistic regression
  - next slide



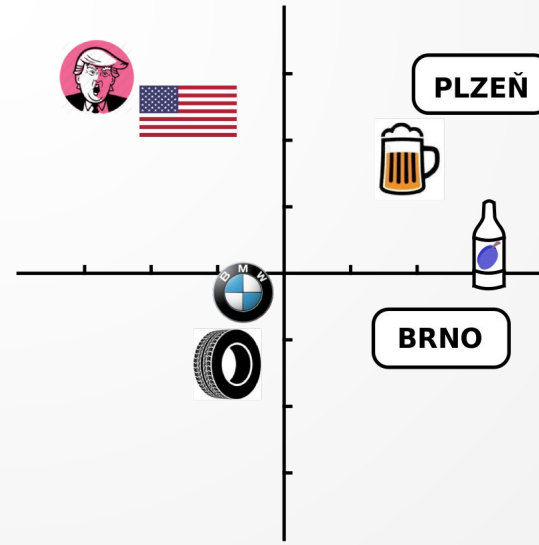
# Deployed algorithms

- Logistic regression - Vowpal Wabbit

**WIDE**

| BRNO  | 0.7 | 2.7 |     | 0.5 |
|-------|-----|-----|-----|-----|
| PLZEŇ | 2.9 | 0.3 |     |     |
|       | 0.3 |     |     | 3.8 |
|       | 0.4 |     | 4.3 |     |

**DEEP**

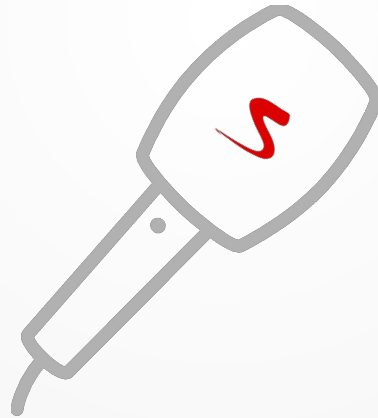


# Deployed algorithms

- Logistic regression - Vowpal Wabbit
  - Reranking of top few hundreds of candidates from CF/CTR-clu
  - many signals/features (ML terminology)
  - user ID, item ID, content tags, vector of interests from targeting department (~hundreds of categories), estimation of age/sex, item IDs user clicked in past ... + combinations of those
  - $\text{score}(u, i) \sim \text{user CTR} + \text{item CTR} + \text{user-tags CTR} + \text{item-age CTR} + \text{itemH-item CTR} + \dots$
  - not enough time to explain in detail
  - intro: [Wide & Deep Learning for Recommender Systems](#)
  - for those who know ML:  
**log features in time of request - add labels later**



# Time for questions



# Diversity / calibration

- Intra List Similarity
  - LDA vectors
  - similar articles have similar vector
  - titles + perex from RSS feed
  - content - from fulltext robot department
  - curse of dimensionality
- Most papers
  - for each position  $p$ :
    - $\arg \max \{ \text{relevancy}(u, i) - \lambda * \text{avg}(\text{similarity}(i, j) \text{ for } j \text{ on position } < p) \}$
  - $O(\text{num\_positions}^2 * \text{num\_candidates})$
  - infinite feed - no research

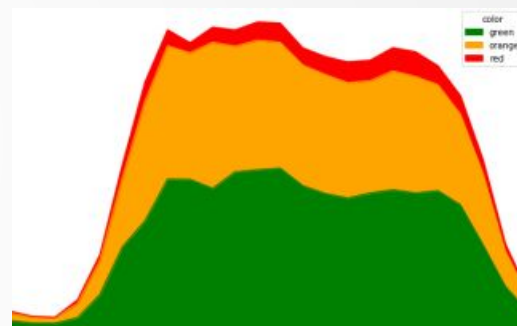
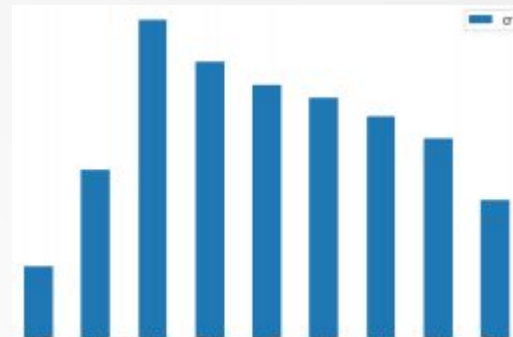


# Diversity / calibration

- Product view
  - feed is supposed to substitute missing boxes on HP (originally)
  - → some degree of low diversity is intended
  - → diversify just in cases if diversity becomes an issue → ILS treshold
  - simple formula:

$$\text{score}(\text{pos}) = 1 / \text{topics\_seen}(\text{pos}) / \text{articles\_topic\_seen}(\text{topic}, \text{pos})$$

- topics - LDA vectors clusters



# Offline evaluation

- always use time-based train-test split
- for regression problems (price prediction)
  - RMSE
- for logistic regression problems (CTR prediction)
  - use LogLoss or RMSE
- for ranking problems
  - use ranking metrics, nDCG, MRR
  - clicks should be high in recommendation list
  - use of RMSE for ranking problem is wrong
    - just multiply score by some constant
    - creators of MovieLens are saying it too



# Offline evaluation

- offline results - weak results
- infinite feed and ranking metrics
- strong position **bias**
- many other **biases**
- Inverse Propensity Score (multi-armed bandits)
  - promising solution
  - many assumptions, limited usability so far
- Offline evaluation of diversity
  - alpha-nDCG = nonsense in practice
  - nDCG-ILS curve?



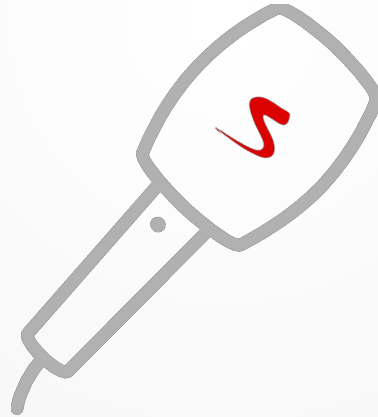


# Online evaluation

- Long-term effects are hard to measure
  - Too good recommender example
  - Complex feedback loops - research needed (simulations)
- Isolation of variants
  - AB variants are affecting each other through training data
  - FB, Netflix, ... - ignoring this issue
  - preparing infrastructure to make experiments possible (log AB variant for each interaction)
  - evaluation of explore strategies will be possible
    - some portion of randomness is beneficial for algorithms
    - how big portion?
  - simulations - promising area of research



# Time for questions



# Research aspects in industry

- “To make great products: do machine learning like the great engineer you are, not like the great machine learning expert you aren’t.” Rules of ML, Google



# Research aspects in industry

- Performance vs. simplicity vs. explainability
- Common scenario:
  - researcher suggest new algorithm
  - 3 days of work = 100 lines of code + 1000 lines of code for proof of concept (offline evaluation)
  - 6 months of work for 4 developers
  - many thousands of lines of production code, 10 components that requires care (monitoring, maintenance)
  - black box for others
- Every nonstationary hyperparameter is problematic
  - never ending story



# RecSys research

- RecSys research is not easily portable
  - different use case - different results for same solution
  - many use cases:
    - Spotify/Netflix - millions of songs, week
    - Seznam.cz - 2000 articles, 50ms
    - Booking.com - lot of domain constraints



# RecSys research

- Almost no academic work is useful in practice:
  - RMSE do not count, still lot of research based on it
  - item-based CF - **Amazon**
  - Matrix factorization - Netflix prize (**Simon Funk's blog!**)
  - ALS - people from **Yahoo**
  - Factorization Machines - **Kaggle** competitions winners
    - **Criteo**
    - **Avazu**
    - **Outbrain**
  - Neural nets, Deep and Wide - **Google, Gravity, YouTube**
  - **Facebook** ... - logistic regression
  - Where is academia? ... practical field indeed



# Last questions



**Thanks**

