

# Companions

Ptáček v Hejnicích 2010

# Companions

- project v FP6
- jistota na stáří
- korpus WoZ dialogů
- TectoMT aplikace

"Petra je sympatický partner pro konverzaci, zajímá se o Vaše fotky a o vzpomínky, které s nimi máte spojené. Je jí kolem třiceti a je to Češka jako poleno, názory na generaci jejích rodičů odpovídají jejímu věku. Má svoje názory, ale už nezná svoji biografii."

# TectoMT aplikace

\$TMT\_ROOT/applications/dialog\_companion/cs/

- Makefile (instalace dat do share)
- scénář tj. seznam bloků:
  - TextSeg segmentace a tokenizace
  - Featurama tagger (neopren)
  - McD parser (pruned.0.02)
  - ZZ t-parser
  - SVM\_ne\_recognizer
  - funktory V.Klimeše
  - Wordnet 2.0 hyperonyma
  - Repair::Set\_node\_schema\_type (pro pmltq)
  - Dialog::DAFRunner

# Dialog::DAFRunner

- dialog manažer - "A co říct teď?"
- DAF = dialog action form - COMIC (Sheffield)
- šablony pro sítě stavů (podle tématů)
- v průběhu dialogu se vyrábějí instance těchto sítí
- aktuální stav pro každého uživatele

přechod:

- testování odchozích hran
- výběr jedné (inverse popularity)
- provedení akce (promluva, práce s fotkama)



# API pro testy a akce

Perlí objekty:

- \$dafrunner
- \$me
- \$you
- \$session
- \$dialog\_history
- \$dialog\_input
- \$dialog\_output

Výstup:

template: "Prima! Vítám Vás," \$you->get\_vocative()

strom:

```
sentence s { nachystat_si => { tak.PREC co.PAT $me.ADDR.perspron you.ACT } }
```

# Klient(i)

Companions 0.3.4 - Jan Ptáček

Soubor Nástroje Zobrazit DEBUG

1\_rodina\_vanoce.jpg [1/12]

System Fiká  
Přemýšlím ...

Historie:  
Uživatel: Dobrý den.

Login Pauza

Adaptace 1/3

czech.companion@gmail.com

Send voicemail Call

Jan: no tak ten maličký tak to je Petr, ten starší to je Jirka no a pak jsme měli ještě nejstaršího máme Milana

czech.companion is online.

czech.companion: Který tam jste?

Sent at 11:36 AM on Monday

Jan: no tak ten maličký tak to je Petr, ten starší to je Jirka no a pak jsme měli ještě nejstaršího máme Milana

czech.companion: Aha. Nevidím Vás tu.

Sent at 12:06 PM on Monday

Jan: no tak ten maličký tak to je Petr ten starší to je Jirka no a pak jsme měli ještě nejstaršího máme Milana

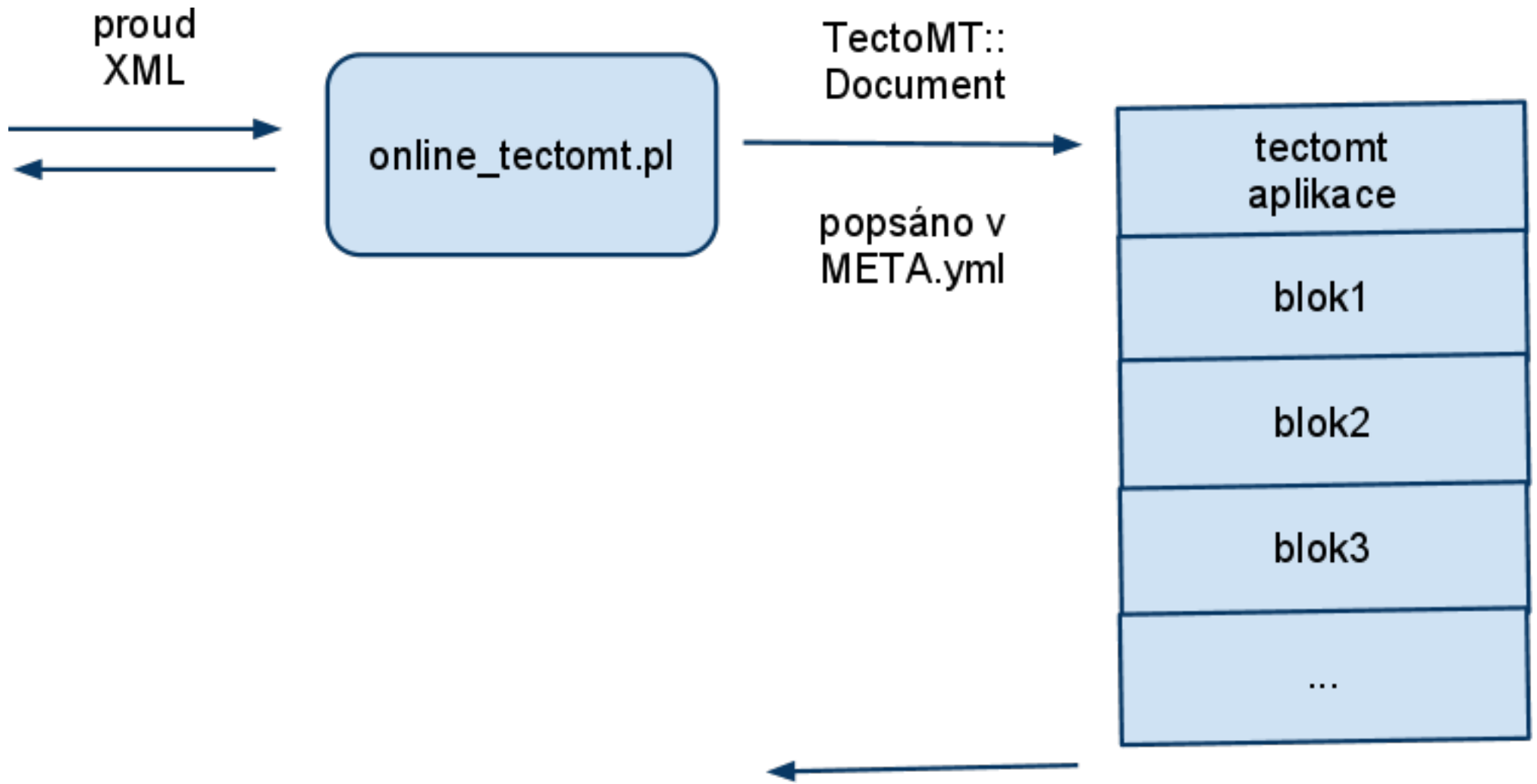
czech.companion: Aha. A který tam jste Vy?

Sent at 1:31 PM on Monday

Jan: .reset

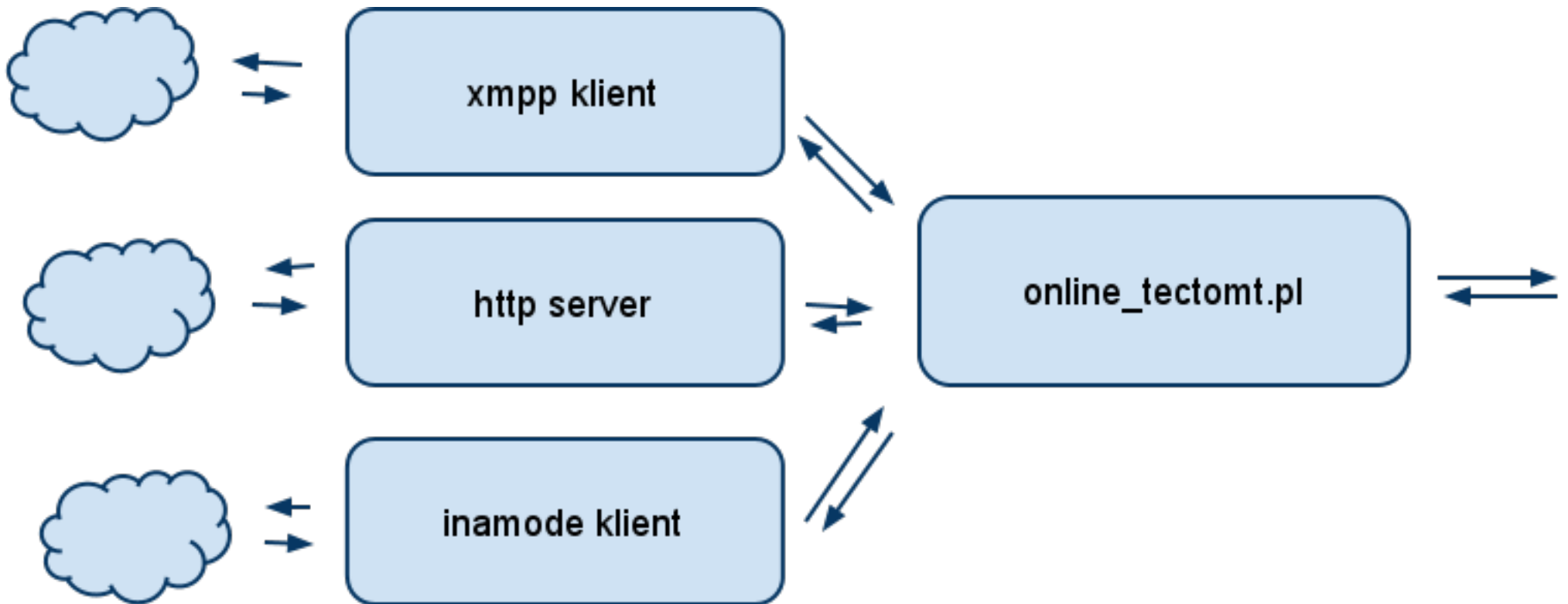
czech.companion: Oukej

# tools/online\_tectomt.pl





# napojení klientů



stačí jen nakonfigurovat aplikaci v jejím META.yml!

# klient z Plzně

## ASR:

- HMM, fine-tuned
- slovník 200 tis. pořad realtime (užíváme 58tis)
- trigram LM
- automatická adaptace na pozadí

## TTS:

- unit selection
- profesionální emotivní řečník
- komunikativní fce (gesto + zásobárna jednotek)  
(SAD\_EMPH)

## Avatar

- Telefonica

# Companions korpus

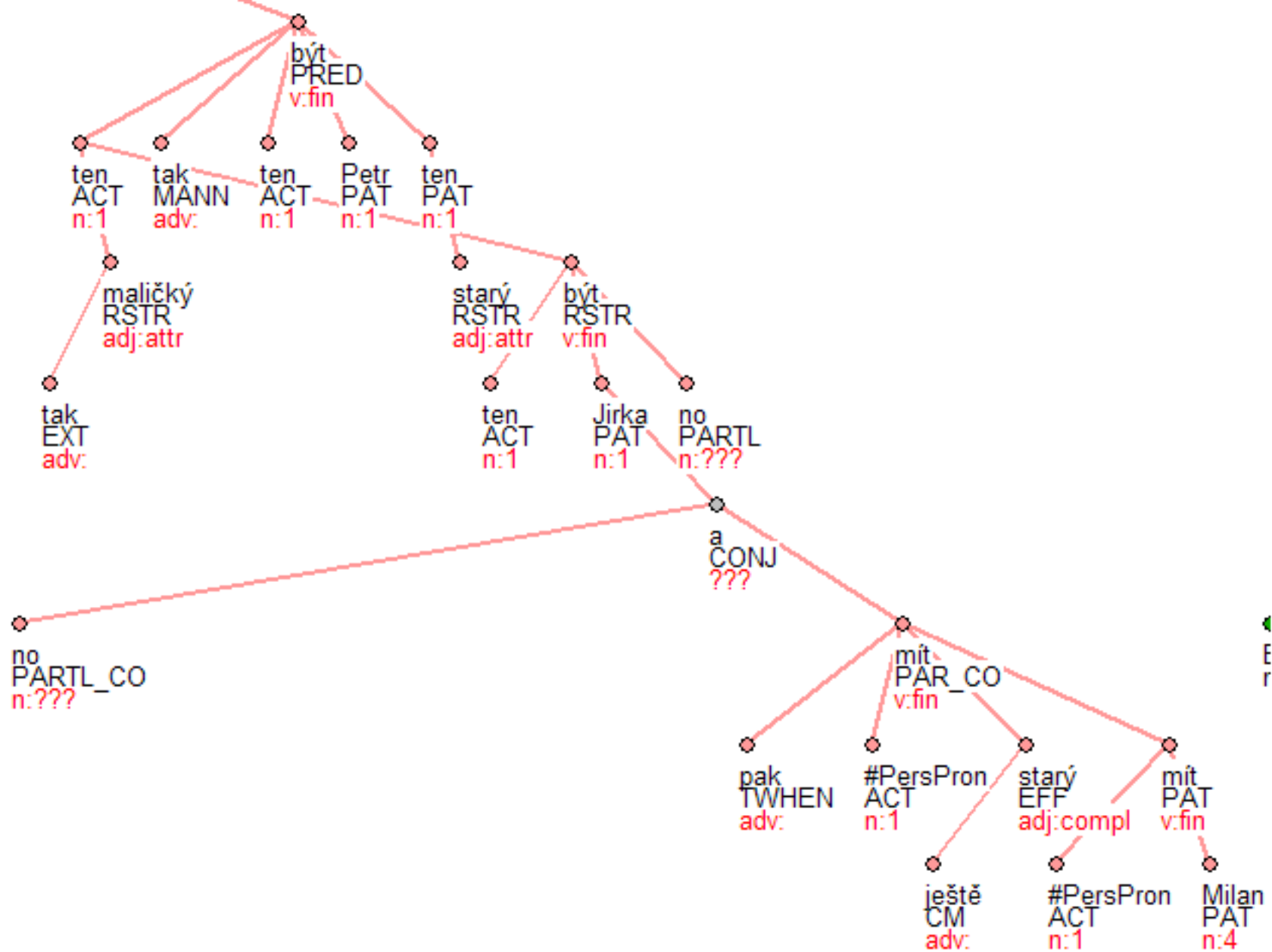
- 100h rozhovorů WoZ
- součást PDTSC, probíhá rekonstrukce
- 29198 vět (27611 vět hotovo)
- 543 tis tokenů
- [https://blackbird.ms.mff.cuni.cz/cgi-bin/pajas/svn\\_file\\_view.pl?details=1&source=companions](https://blackbird.ms.mff.cuni.cz/cgi-bin/pajas/svn_file_view.pl?details=1&source=companions)

## anotace dialogových aktů

- Silvie, upravená podmnožina DAMSL tagsetu
- pilot 1000 vět, da-tagger:
  - 64% baseline
  - 69.3% +- 0.9%
- ostrá anotace:
  - ~6tis výpovědí, hotová

*demo*

*jabber: czech.companion@gmail.com*



**me:** Tak to je maličky, starý Petr, který je Jirka a který my pak máme ještě staršího, máme ho tak, že který je Jirka, takového, že my máme Milana.

**Pavel:** Tyy vole!!! :) A z toho budeš schopný něco

no tak ten maličký tak to je Petr, ten starší to je Jirka no a pak jsme měli ještě nejstaršího máme Milana

