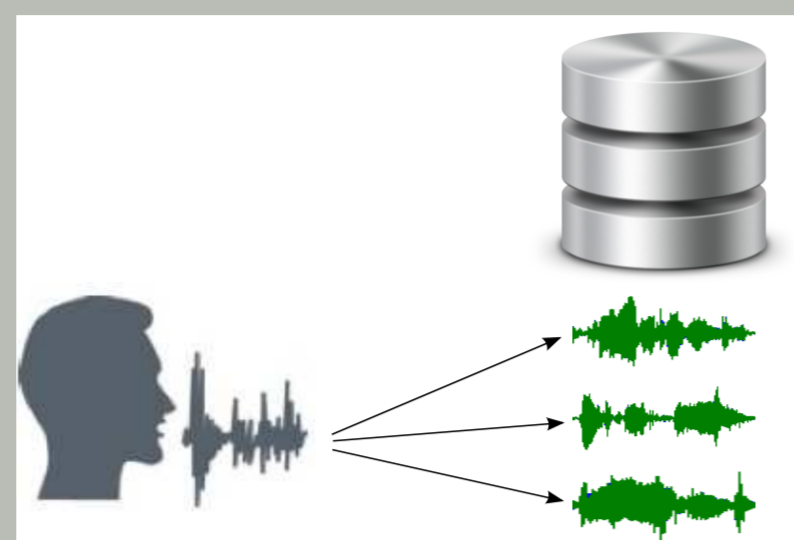


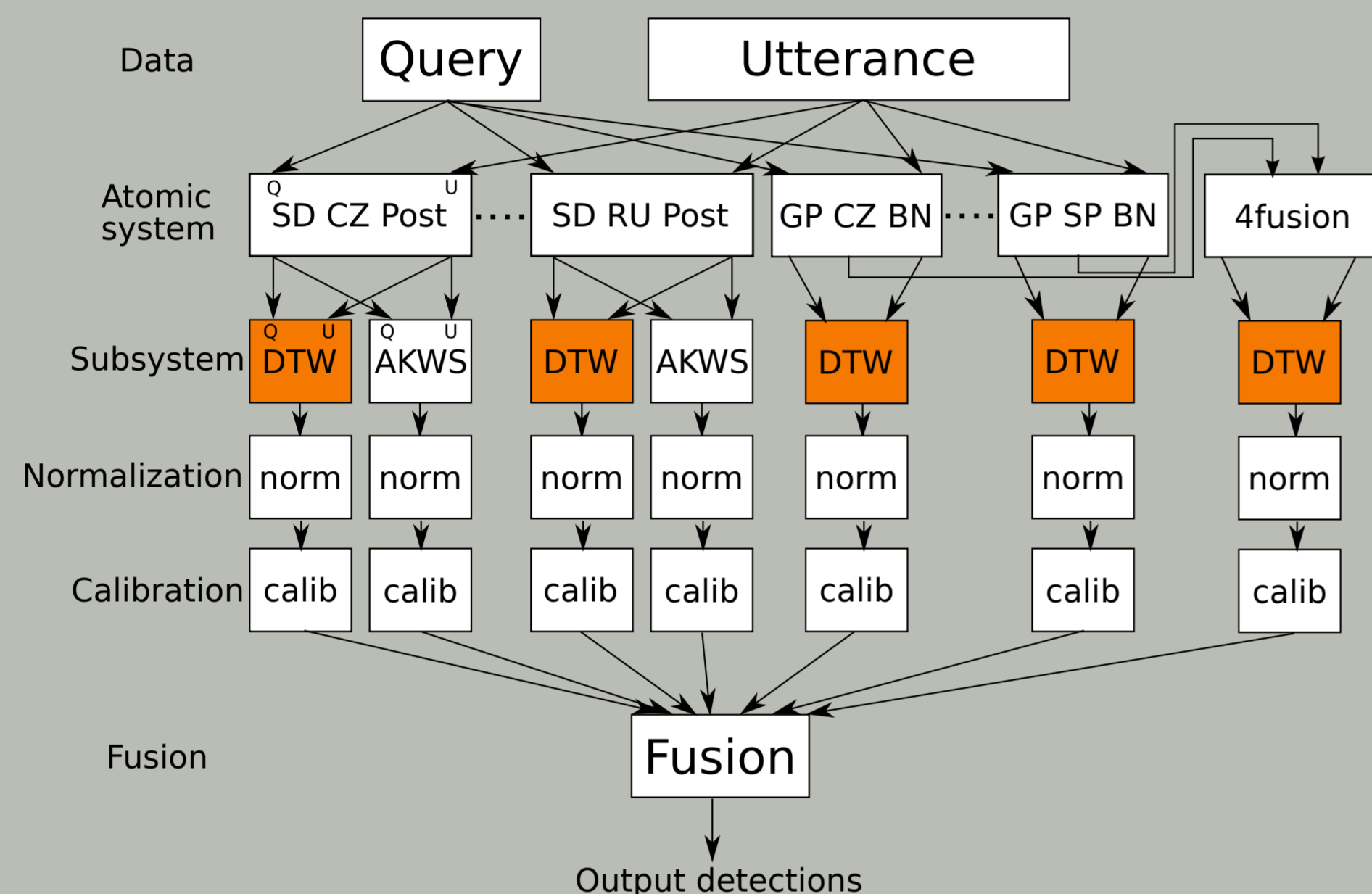
Úvod

Navržený systém slouží k detekci uživatelem specifikovaných řečových frází v rozsáhlé databázi s omezenou znalostí cílového jazyka. Pro detekci byla použita metoda dynamického borcení času. Systém byl úspěšně použit pro evaluace QUESST.



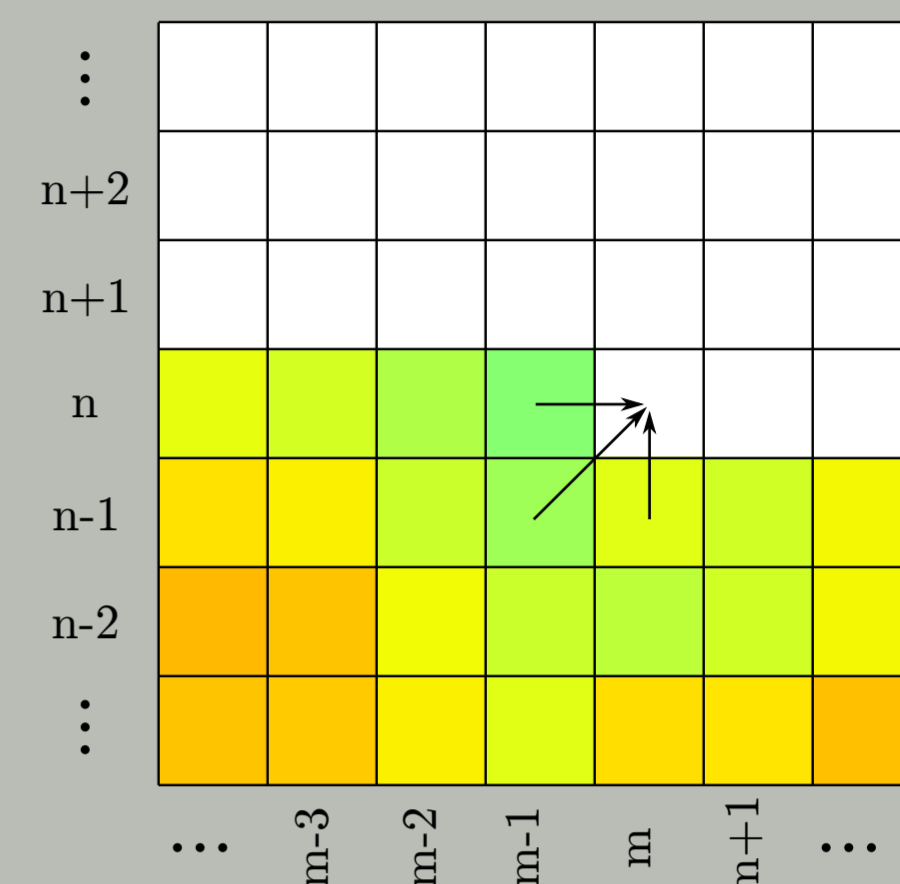
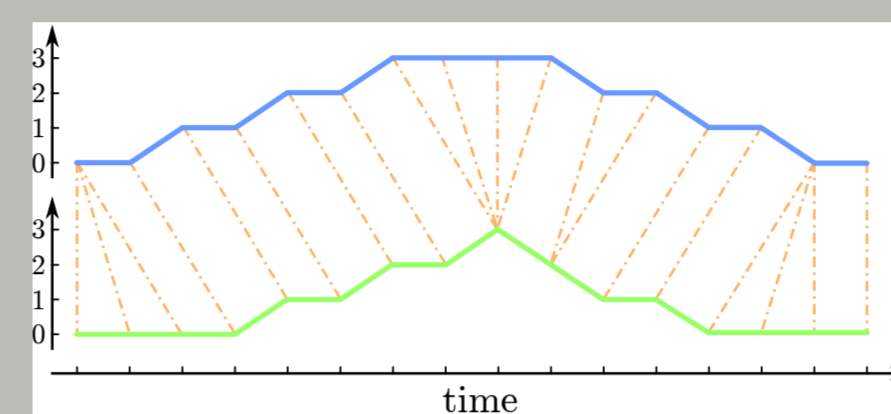
Detekce klíčových frází (QbE)

Systém pro detekci klíčových frází (Query-by-Example, QbE). Vstupem je hledaný výraz (*query*) ve formě řečové nahrávky, který se vyhledává v databázi s promluvy (*utterance*). Ze vstupních audio nahrávek jsou extrahovány důležité příznaky (vektory) pro další zpracování. Autor pracoval na podsystémech DTW (oranžově označeno).



Dynamické borcení času (DTW)

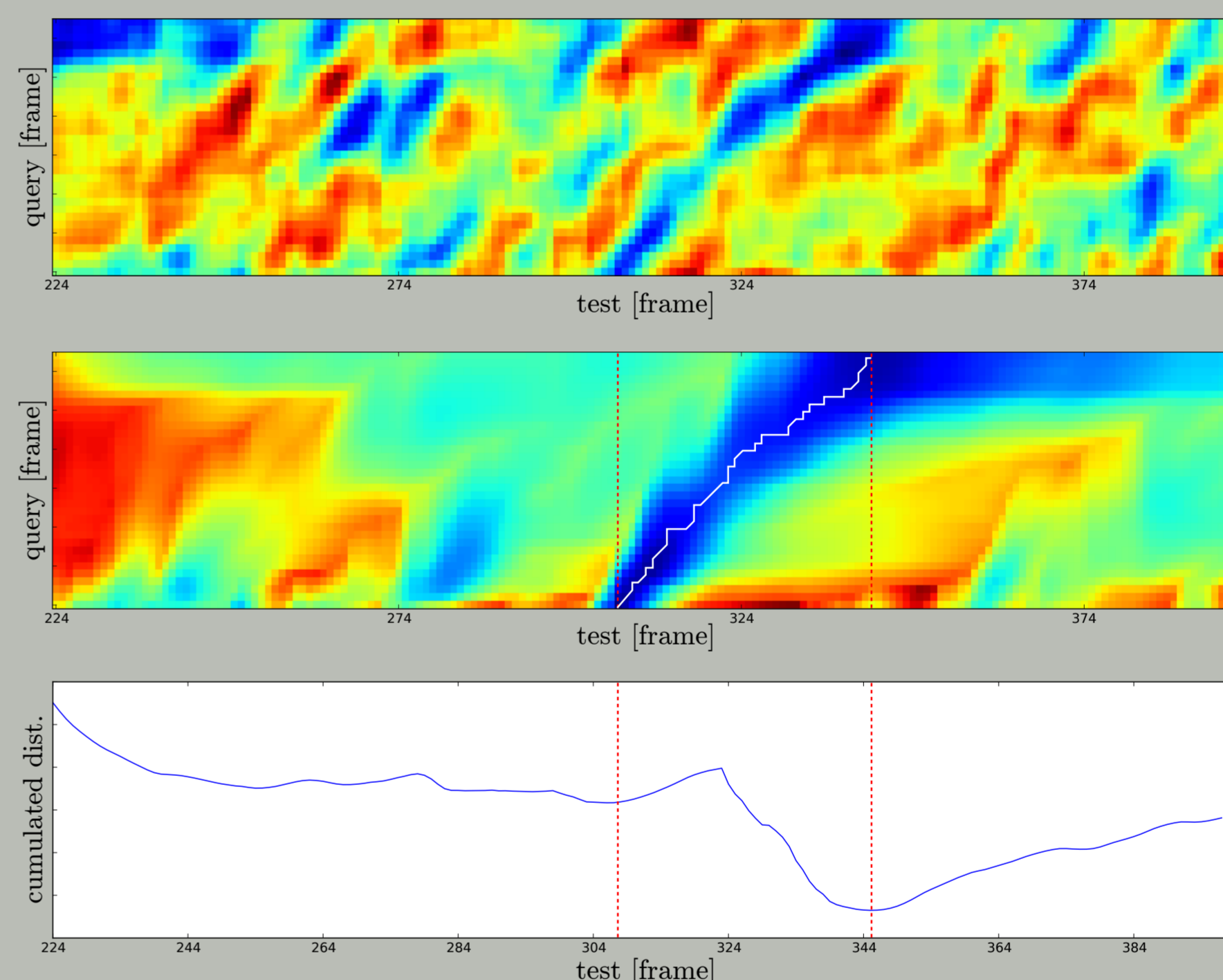
Dynamické borcení času (Dynamic Time Warping, DTW) je metoda vhodná pro detekci izolovaných slov, která porovnává dvě náhravky řeči a snaží se je optimálně zarovnat pomocí borcení časové osy.



Pro nalezení optimálního zarovnání je potřeba vypočítat **distanční matici**, která obsahuje vzdálenosti mezi vstupními vektory. Dále je potřeba **kumulační matice**, která postupně kumuluje hodnoty z distanční matice podle uvedeného vzorce níže. Váhy w_d , w_h , w_v při hledání optimální cesty byly nastaveny na hodnotu 1.

$$C(n, m) = \min \begin{cases} C(n-1, m-1) + w_d \cdot D(n, m) \\ C(n-1, 1) + w_h \cdot D(n, m) \\ C(n, m-1) + w_v \cdot D(n, m) \end{cases}$$

Ukázka matic potřebných pro detekci požadovaného slova v testovací promluvě. Profil zkreslení slouží k určení koncového bodu cesty.



Metriky pro měření vzdálenosti vektorů

Kosinová vzdálenost:

$$d_{\cos}(\mathbf{u}, \mathbf{q}) = 1 - \frac{\mathbf{u} \cdot \mathbf{q}}{|\mathbf{u}| \cdot |\mathbf{q}|}$$

Pearsonova korelace:

$$d_{\text{corr}}(\mathbf{u}, \mathbf{q}) = 1 - \frac{(\mathbf{u} - \bar{\mathbf{u}}) \cdot (\mathbf{q} - \bar{\mathbf{q}})}{|\mathbf{u} - \bar{\mathbf{u}}| \cdot |\mathbf{q} - \bar{\mathbf{q}}|}$$

Euklidovská vzdálenost:

$$d_{\text{euc}}(\mathbf{u}, \mathbf{q}) = \sqrt{\sum_{k=1}^L (u_k - q_k)^2}$$

Logaritmus věrohodnosti na základě kosinové vzdálenosti:

$$d_{\log\cos}(\mathbf{u}, \mathbf{q}) = -\log\left(\frac{\mathbf{u} \cdot \mathbf{q}}{|\mathbf{u}| \cdot |\mathbf{q}|}\right)$$

Logaritmus věrohodnosti na základě skalárního součinu:

$$d_{\log\text{dot}}(\mathbf{u}, \mathbf{q}) = -\log(\mathbf{u} \cdot \mathbf{q}),$$

Reference

- ▶ Igor Szöke, Miroslav Skácel, Lukáš Burget and Jan Černocký. Coping with Channel Mismatch in Query-by-Example - BUT QUESST 2014. Accepted at ICASSP 2015, 2015.
- ▶ Stan Salvador and Philip Chan. Toward Accurate Dynamic Time Warping in Linear Time and Space. *Intell. Data Anal.*, 11(5):561580, October 2007.
- ▶ Meinard Müller. *Information Retrieval for Music and Motion*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

Vyhodnocení

Jako nejvíce robustní metrika pro měření vzdálenosti mezi vstupními vektory během testování byla zvolena Pearsonova korelace d_{corr} , která byla použita i pro evaluační data.

Systém	C_{nxe} / C_{nxe}^{\min}
BUT 4fusion	0.473 / 0.466
BUT GP CZ BN	0.536 / 0.528
NTU-NPU-I2R	0.602 / 0.598
EHU	0.621 / 0.599
SPL-IT	0.659 / 0.508
CUHK	0.683 / 0.659
IIIT-H	0.921 / 0.812

Závěr

Navržený systém překonal všechny ostatní účastníky soutěže QUESST na konferenci MediaEval 2014. Lepších výsledků dosáhl jen systém složený z fúze 4 atomických systémů na bázi bottlenecků příznaků. Porovnání systémů ostatních účastníků je uvedeno v předchozí tabulce.