LIA-Thales at Duc'06
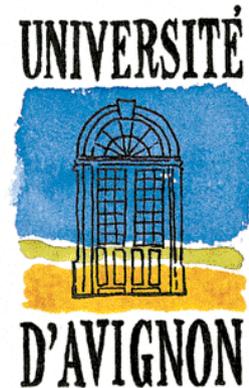
A Fusion strategy for summarization using multiple systems

benoit.favre@univ-avignon.fr
Laboratoire d'Informatique d'Avignon (LIA)
Thales Land & Joint Systems, France

Thanks to: Frederic Bechet, Patrice Bellot, Florian Boudin, Marc El-Beze, Laurent Gillard, Guy Lapalme, Juan-Manuel Torres-Moreno

Outline

- Why apply output fusion to summarization ?

- Short description of the 5 systems

- The Fusion Strategy

- DUC results

- Analysis and conclusion

Why apply fusion to summarization ?

- Performance variation of the systems on the same topic

- Successful in other domains
  - Classification
  - Speech recognition (ROVER)
  - Speaker identification

- Robustness of hyper parameters estimation
  - Small training dataset

Principle
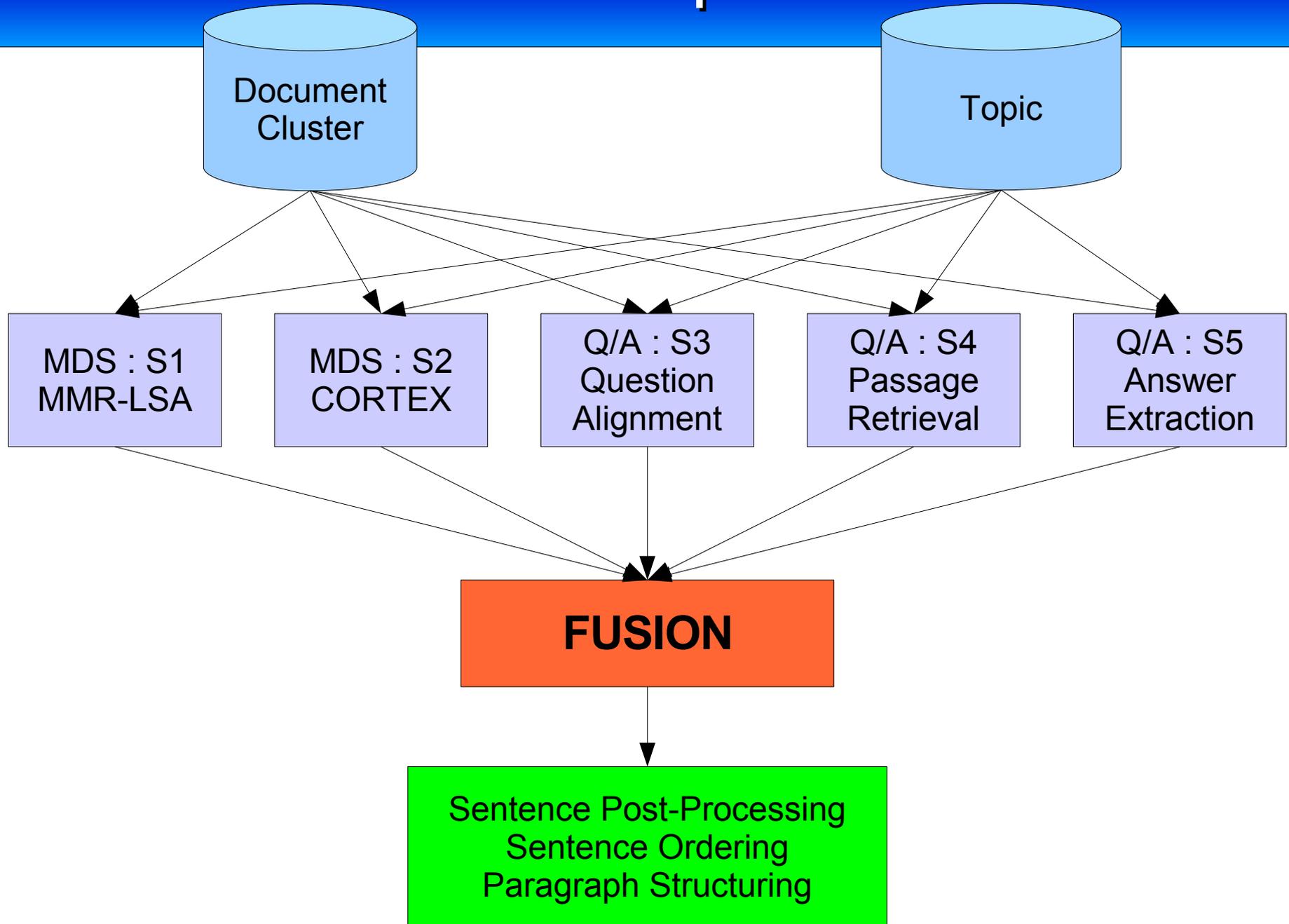
The MDS systems

- MMR-LSA (system S1)
  - Reduced cooccurrence space representation
  - anti-redundancy, topic centered ranking
  - greedy optimization
- CORTEX (system S2)
  - Feature based sentence ranking
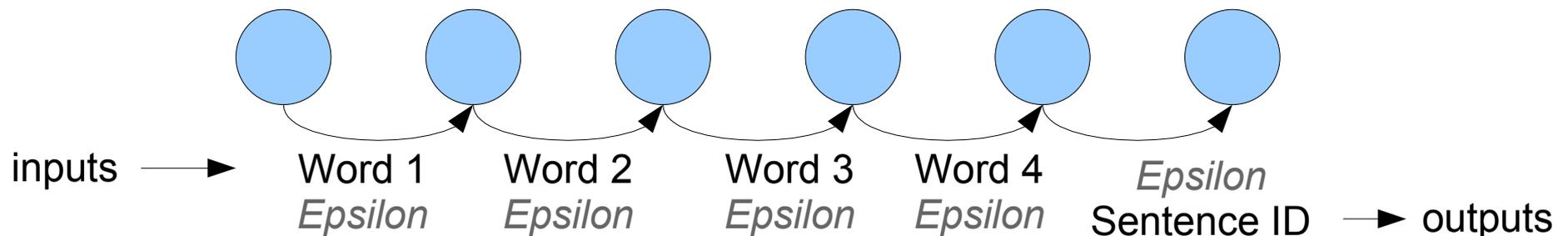 - Elaborate decision process

The Q/A systems

- Question modeling (system S3)
    - N-term with variable length gap model, sentence position and vocabulary coverage.

- Parts of the LIA Q/A system
    - Passage retrieval (system S4)
        - Question density in the context sentences + expected answer type (when possible)
    - Answer extraction (system S5)
        - "Compacity" score around the supposed answer

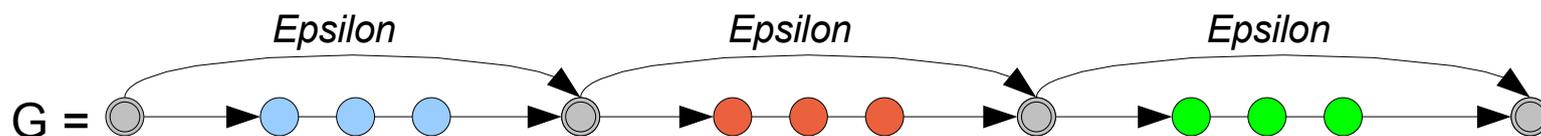Not tuned on the development corpus (DUC'05)

The Fusion Strategy

- The idea :
  - Systems output : ranked sentence lists.
  - All sentences are represented as a Weighted Finite State Transducer (WFST).
  - Output summary : the best path in this WFST

- A sentence as a FST :

inputs → Word 1 / *Epsilon* → Word 2 / *Epsilon* → Word 3 / *Epsilon* → Word 4 / *Epsilon* → *Epsilon* / Sentence ID → outputs
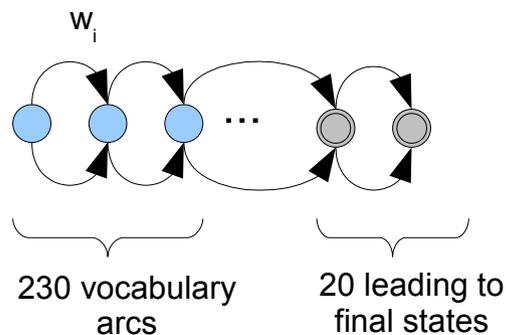
Building the Hypothesis Space

- Graph of the 30-best sentences of each system

- Only paths containing around 250 words are kept for the scoring process

$\oplus$ G = Unweighted hypothesis space

230 vocabulary arcs

20 leading to final states

Weighting the hypothesis space

- At Final state level : distance to 250 words

- At Sentence level
  - Number of systems that voted for the sentence
  - Best rank of the sentence in the lists

- At Word level
  - Heuristics to reduce pronominal anaphora and relative temporal references

- All costs are tuned on DUC'05 to maximize ROUGE scores

Best Path :

5 votes

• president Boris Yeltsin announced that Russian forces would halt their offensive in rebel Chechnya from midnight and start a partial withdrawal from some areas , but warned that Chechen separatist fighters would be hit hard if they stage terrorist attacks .

4 votes

• the latest report shows that the casualties of the Russian joint troops in Chechnya in a series of terrorist blasts made by Chechen rebels Sunday night have reached over 100 , said the Russian Interior Ministry on Monday .
• preliminary reports indicate that 36 Russian servicemen were killed and another 74 injured in five terrorist attacks in Chechnya on Sunday , Russian presidential aide on Chechnya Sergei Yastrzhembsky announced here Monday .
• Chechen rebels committed five terrorist blasts in several key cities of the war torn republic Sunday night , killing 36 Russian servicemen and wounding 74 , Russian presidential aide on Chechnya Sergei Yastrzhembsky announced .
• russian federal troops continued aerial and artillery attacks on rebel positions Friday in the breakaway republic of Chechnya as Moscow declared victory in the first stage of offensives against Islamic militants there .

• about 50 persons have been detained in the breakaway republic of Chechnya , eight of whom may have taken part in the suicide bombing in the city of Argun , 10 kilometers southeast of the Chechen capital of Grozny , Russian presidential aide Sergei Yastrzhembsky on Chechnya told an interview in Moscow Wednesday .
• by Wednesday , Russia have detained more than 200 people in Chechnya and its neighboring republic of Dagestan on suspicions of involving in the terrorist blasts Sunday night , which caused over 100 person casualties to Russian troops .
• russian federal forces continued air raids on rebel positions in Chechnya Wednesday as restrictions were imposed in the breakaway republic to prevent rebel attacks during the May holidays .
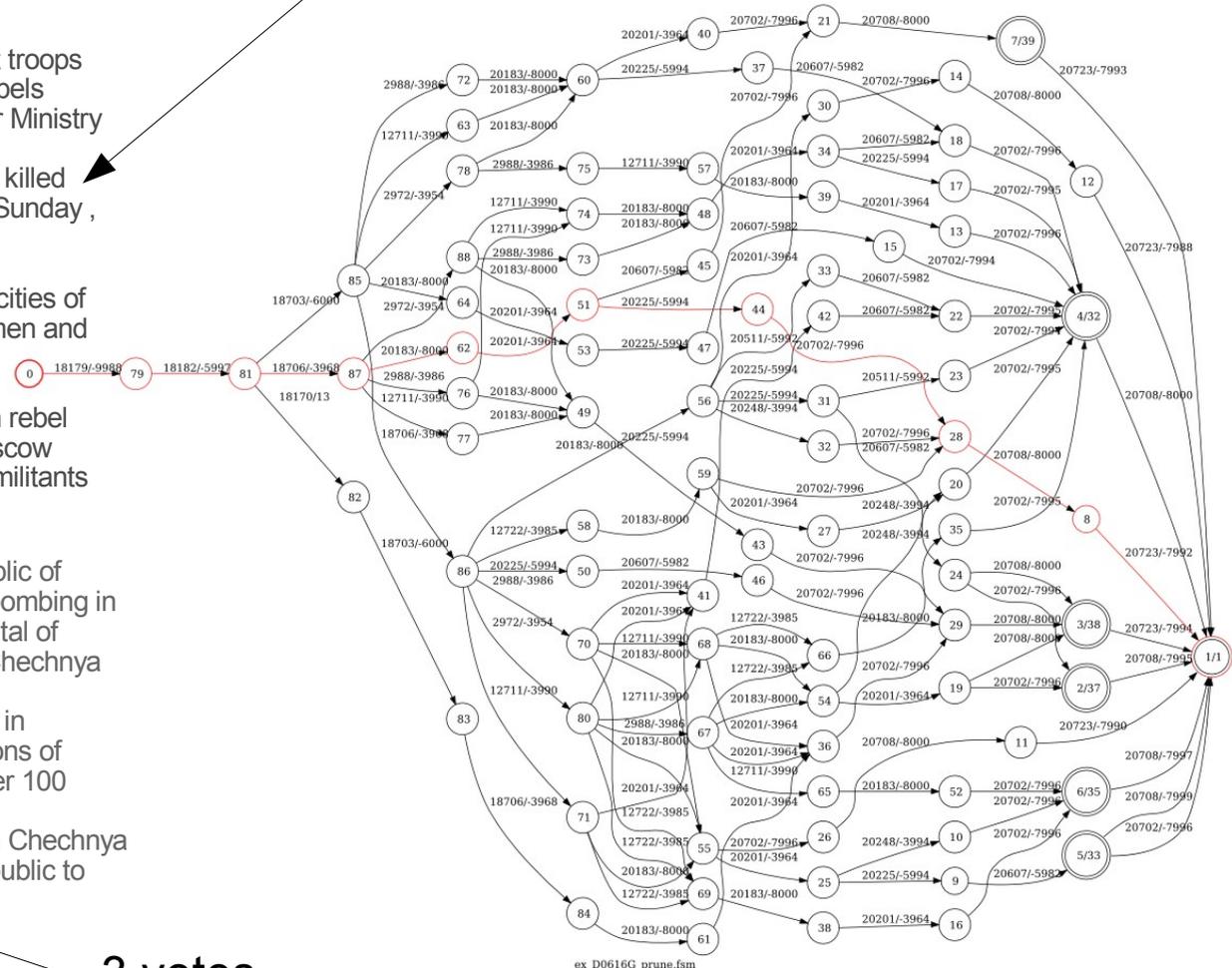
3 votes
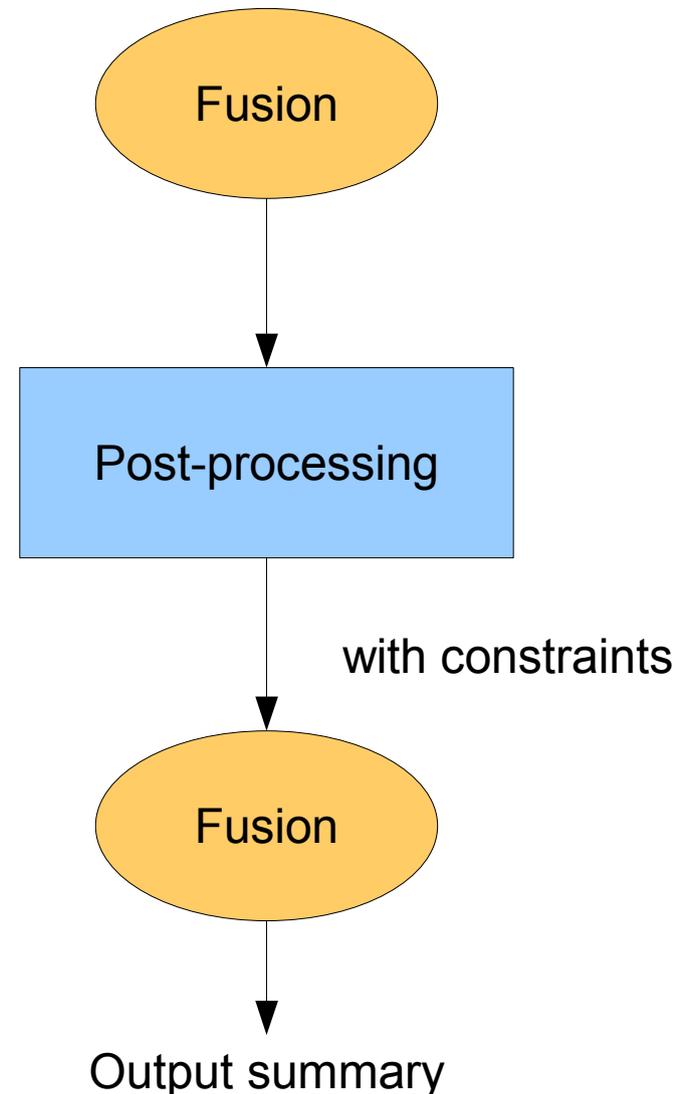
Post-processing

- **Change length**
  - Acronym and person name rewriting
  - Link words and say clause removal
  - ...
- **Sentence reordering**
- **Paragraph structuring**

Fusion

Post-processing

with constraints

Fusion

Output summary

Behavior of the Fusion

- ## Distribution of votes

vote	2 syst.	3 syst.	4 syst.	5 syst
2005	100%	72.7%	35.6%	07.4%
2006	100%	79.4%	44.5%	12.4%

Predicts comparable performances

- ## Systems weight

System	S1	S2	S3	S4	S5
2005	15.0%	20.7%	23.0%	17.4%	23.9%
2006	16.5%	21.2%	22.8%	16.8%	22.7%

Automatic evaluation : DUC results

- ## LIA-Thales Rank over 34 systems (our id=28)

Manual scores	rank
Grammaticality	7
Non-redundancy	31
Referential clarity	6
Focus	13
Structure and coher.	19
Ling Quality Mean	14
Resp-Content	8
Resp-Overall	8

Automatic Scores	Rank	Better	Worse
		significantly	
Pyramid	6		
Rouge 2	6	2	25
Rouge SU4	5	1	26
BE	6	2	26

- ## Bad non-redundancy

  - limited capture of redundancy in ROUGE/BE

  - voting tend to force redundancy

Fusion : conclusions

- Advantages :
  - Outperforms the best system (ROUGE)
  - Straightforward
  - Toolkits available (we use the AT&T FSM toolkit)
  - Flexible (support for hand crafted heuristics)

- Drawbacks :
  - Independent selection of sentences
  - Parameter tuning using a development corpus

What's next ?

- Incorporate anti-redundancy in the fusion process

    – Pre-process fusion inputs

    – Iterative WFST reweighing

    – Summary selection (genetic algorithm)

- Select sub-sentence components

    – sentence reduction as a pre-processing : generate alternate sentences as input of the systems

    – not included in the submission, will be presented in a future article

Thank you

benoit.favre@univ-avignon.fr

Co-authors : Frederic Bechet, Patrice Bellot, Florian Boudin, Marc El-Beze, Laurent Gillard, Guy Lapalme, Juan-Manuel Torres-Moreno

Short Description of the 5 systems

- Systems developed for the Question Answering (Q/A) task

  - Topics contain factoid sub-questions (who...)

- Multi-document summarization (MDS) systems

  - The usual way to participate in DUC

- « Optimized on DUC'05 » *vs* « Off the shelf »

  - Validate the utility of training data

  - Prevent overfitting

- ## Post-processing

  - ### Acronym and Person name rewriting

    "Attention deficit hyperactivity disorder (ADHD) ... a consensus statement on ADHD and its treatment ... all the criteria of ADHD ..."

    "Bill Clinton ... Clinton ... Clinton ..."

  - ### Link words removal, say clauses removal

    ~~Moreover,~~ the president is ...

    ... ~~said the judge~~.

  - ### Cleanup punctuation

  - ### ...

Other components of our submission (2)

- ## Sentence ordering
  - Topic classification
    - Temporal or Geographic dimension
    - Specific or General
  - Partial order, Temporal first or Geographical first
  - Introduce paragraphs with Year or Country Name

    "In 1999, ..." , *"In Brazil, ..."* (if Specific)

- ## Paragraph structure
  - Cut at Year/Country change

Automatic evaluation : comments

- It's worth optimizing on dev data
    - same trend of optimized vs not optimized systems on 2005 and 2006 data

- Fusion : a gain over the best one
    - F1 and F2 have better ROUGE scores than S1

- Robustness to overfitting
    - F2 (5 systems) better than F1 (3 best systems)