

**DUC-05 - DUC-07**

**ROAD MAP PROPOSALS**

**WORKING GROUP :**

**Karen Sparck Jones, Hans van Halteren,  
Marie-Francine Moens, Guy Lapalme,  
Dragomir Radev, Bonnie Dorr,  
Paul Over, Ed Hovy,  
Kathy McKeown, Donna Harman**

- 1. Background - DUCs 1-4**
- 2. WG Topics**
- 3. RM Proposal and Followup**

+

## 1. Background

( Detailed notes to be posted on DUC-04 website )

**Inputs :**

**Outcomes of post DUC-03 Specific Working Groups**

**Original Road Map**

**DUC 01-04 Specifications and Overviews (thanks, Paul)**

## Post DUC-03 Specific Working Groups :

1. Improving quality questions  
(A Nenkova, R Passonneau)
  - => new question set
  - => evaluation study for multiple summaries
  
2. Extrinsic evaluation of headline summaries  
(B Dorr, R Schwartz)
  - => partially done
  
3. Development and assessment of ROUGE  
(CY Lin)
  - => new versions, documentation
  - => studies of DUC results, ROUGE behaviour  
[ summaries etc to Web Site ]

Road Map Working Group :

Original Road Map 2000

Over 4 years :

develop corpora, evaluation methods

progress from

extract to abstract

single doc to multidoc

simple genre to complex genre

English to other language

plain to eg evolving, answering

intrinsic to extrinsic evaluation

## Achievements :

developed evaluation methodologies, tools

- qualitative, quantitative measures
- sets of training/test data

conducted careful comparative experiments

- variations of summarising task (needs, lengths)
- alternative strategies and tactics

developed extractive summarising ideas, systems

- single document, multidocument

reached better than rockbaseline performance

(built some community and confidence)

## Limitations :

quantitative measures crude, extraction oriented

data all news material

task variations rather limited and unstringent

little single-doc, more multi-doc summary

evaluation intrinsic, or weak extrinsic simulation

all extractive type summarising, no abstracting



Conclusions on DUCs 1-4 :

if it's reflective, extractive, multi-document  
news summarising you want

==> we've been there, done that

**BUT**

**LOTS MORE TO DO**

How tackle ?

Working Group deliberations (thanks everyone) :

accepted need to move

from only news text  
to abstracting  
for solid need  
with extrinsic evaluation

discussion focused on genre, with implications  
for summary type, user need, system performance

but suggestions narrow, orthogonal -

eg transcribed speech headlines

eg book fiction resume

eg legal report collocation

too hard, too dissipative, ...

what to do for synergy, progress ?

## 2. Proposals

==> to all-singing all-dancing solution -

**THE DUC MODEL 2 ROADMAP ==>**

The Model :



Summary context :  
a professional information analyst

Summary task :  
resume of information on an urgent situation

Summary data :  
multiple types of source

Summary evaluation :  
intrinsic, biased to situation  
ie simulated extrinsic  
pseudo-real extrinsic  
  
component level, whole-system level

**EXAMPLE - trigger situation :**

event -

huge volcanic explosion in remote Concarpia

user -

head of emergency relief service

data -

news, geographic data, geological information,  
public health/disease data, emails ...

evaluation -

pertinent information, well organised  
per data type, overall

EXAMPLE - output resume :

Major volcanic explosion Mt Popup, Concarpia	< news
at mapref PQR,	< geog data
with apparent widespread damage.	< news
Nearest towns are Mrketplce and Trtop	< gazetteer
with combined populations of 50,000.	< encycl
The nearest railway is at Bigtwn, 50 miles away,	< encycl
the roads are mere dirt and the nearest airstrip is	< encycl
250 miles away at Capita.	< encycl
RedCross in Capita is signalling asking for help	< voice/email
and Euroemergaid is suggesting coordinated response.	< email
It's the rainy season so Jipkin's disease outbreaks	< sci literature
is especially likely.	< sci literature
The most useful drugs for this are Prevcyclene	< sci literature
and gin.	< sci reports
We have no Prevcyclene in stock.	< org records
Bigpharma in France have a reserve supply.	< emerg database



**This sort of thing is not frivolous -  
plenty of people will want to have such capabilities :**

Eco Emergencies - storms, floods, ... (The Guardian 28.4.04)

Red Cross/Red Crescent Study:

natural disasters trebled last 3 decades

numbers killed fallen but numbers affected annually  
quadrupled

Insurance companies:

economic and insurance losses doubling every decade

European heatwave 2003 deaths:

lack of information, coordination exacerbated problems

## Model features :

new challenge, step up from DUC 01-04

real world relevance

hospitable to specific interests

allows levels of participation

attractive to new players

feasible data provision

manageable (decomposable, staged) evaluation

ie \* something for everyone \*

component/overall level activities :

summarise within/across data type -  
genre, multi/single doc ...

explore strategy range -  
extract/abstract ...

investigate presentation aspect -  
segmented/integrated ...

.....

Issues :

what generic domain ?

what task specification sequence ?

what evaluation specification sequence ?

need some selection for control

manageable effort

informative comparison

Strategy (behind proposal, for specification) :

**\*many\*** factors affecting summarising -

input factors

(source form, links, language .. )

output factors (ie choices)

(summary format, register, ...)

cannot just pick 'n mix on input and output features

output guided by summary **\*purpose\***

(use, audience, ...)

under constraints from input features

==>

therefore :

adopt generic purpose ie situation wanting summaries  
\*as suggested\*

refine for series of manageable versions  
use to constrain sets of input conditions  
apply to choose output specifications

eg

informative resumes in flooding emergencies  
for administrators / for police  
inputs with properties P,Q,R / P,S,T  
outputs with features D,E / F,G,H

Slogan :

careful, detailed elaboration of basic suggestion

so everyone is reasonably satisfied

so, folks, let's get stuck in ...

( can assemble material, reference data )

Discussion on RM 04 at DUC 04 - Main points :

Issues :

1 generic approach

- build on experience (no global restart), so allow for current concerns (eg speech, headlines) expand (from news) to other source types involve volunteer subgroups

2 topic (situation) design (and data collection)

- check for realism explore generic/specific tradeoff focus on summarising NOT just IE/QA



### 3 output scenario

- layer model (drill down)

initial 'one time' output

top bullets/exec summary

lower bullets/resume

source type/information type

output formatted by type/integrated

later query-driven, interactive

### 4 specialised domain knowledge

- amount needed ? (shouldnt overwhelm)

Procedure :

by ACL Barcelona July 2004 :

have generic domain

example situation topics (consult)

trials for data supply (Web crawls)

(different types must connect with  
topic eg temporally)

==> offer more specific RM proposal

by Fall 04

have DUC 2005 specification

- evaluation methodology !

(NIST-led studies with with community feedback)

Particular points :

accommodating summarising for

1 input variations -

single doc and multidoc

one language and across language

2 output conditions -

very short to quite long output

focused or open content

tight or loose structure

??

(possibly also use input metadata,  
allow output graphics ....)

deal with

1 input variation

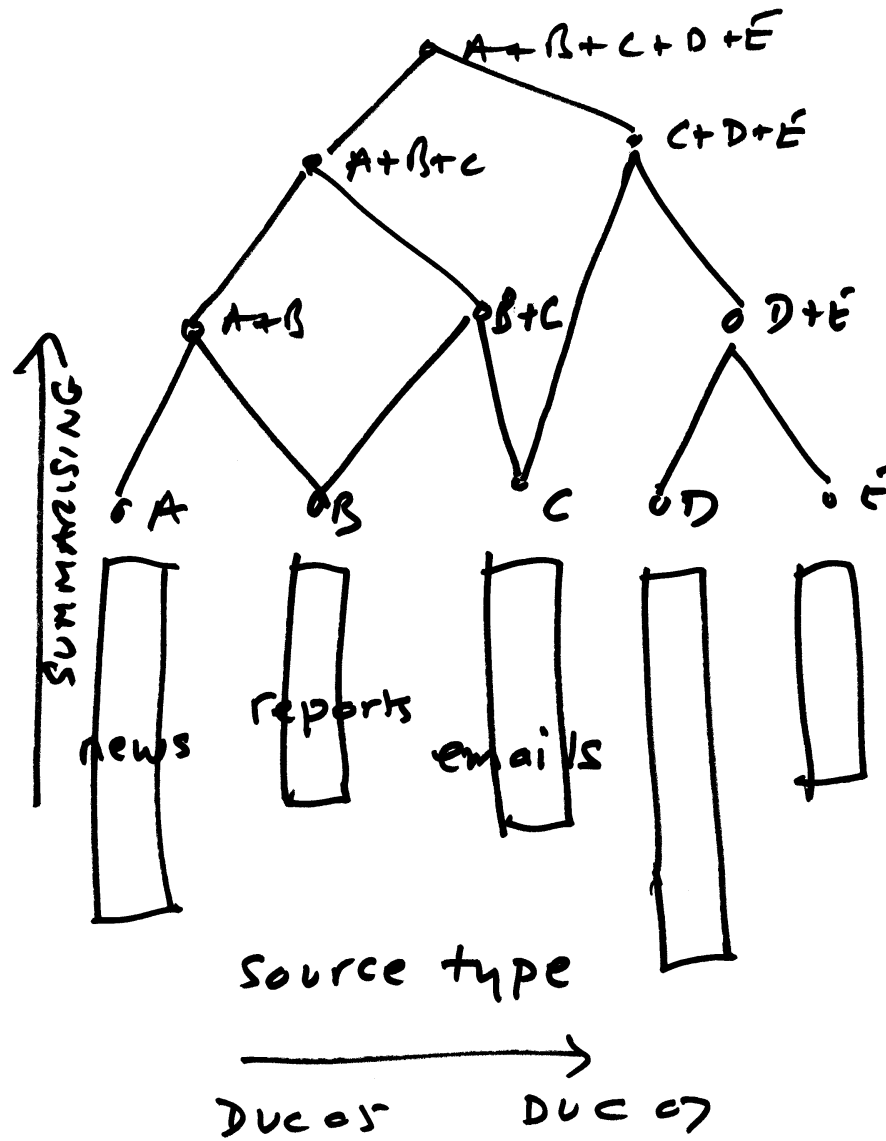
by data driving

ie systems respond to what they find  
in various source types supplied  
(not all for all topics)

2 output conditions

by natural response to guidance for  
topic situations of different kinds  
layers in summarising 'tree'

# THE ROADMAP LAYER TREE



DUC 05 - 07 natural evolution -

summarise for situation topics

eg on new SARS outbreak for WHO

progress from familiar source types eg news

towards new eg reports, emails

(A, B, C in figure)

summarise over single type and type combinations

eg report, report+emails

(B, B+C)

produce brief and fuller summaries for any level

eg oneliner for report and para for report,

oneliner and halpage for report+emails

evaluation problems :

intrinsic evaluation -

quality questions OK

use of ROUGE when fixed length summaries

unrealistic ?

extrinsic evaluation -

'responsiveness' notion plausible

but viable detailed implementation

with complex situations ?

investigation for DUC 05 specification

slogan for DUC 05 - DUC 07 :

SUMMARISING AS INFORMATION FUSION

FOR A USER SITUATION