Appendix

Author: Kavishwar Wagholikar
Date: 18 Feb 2012
Email: wagholikar.kavishwar@mayo.edu

This document provides a detailed description of the rulebases for the text processor and for the guideline reported in the paper titled "**Clinical Decision Support with Automated Text Processing for Cervical Cancer Screening**"

Sections A and B, provide a detailed description of the rulebases. Section C refers to the Drools implementation of the rulebases, that the users can adapt and re-use to develop their systems.

A) Detailed overview of the text-processor rulebase

The text-processor rulebase consist of lexer and parser rules.

Lexer rules map the text pattern to concept code. The following gives the list of text/regular expressions for concept abbreviations/codes:

Abbreviation/ Concept Code	**Text pattern**
GECA	glandular epithelial cell abnormality
PHPV	hpv testing was performed by digene hybrid capture ii and is positive for one or more of hpv types [\d|\s]+ and
NHPV	hpv testing was performed by digene hybrid capture ii and is negative for hpv types [\d|\s]+ and
ASCH	atypical squamous cells cannot exclude high grade squamous intraepithelial lesion
NOEC	inadequate endocervical transformation zone component
ASKS	atypical squamous cells of unde[\w]*rmined significance
HPVQNS	hpv testing cancelled due to insufficient quantity of testable material
SCC	consistent with invasive squamous cell carcinoma
NIL	negative for intraepithelial lesion or malignancy
DNOEC	an inadequate endocervical transformational zone component is not

	necessarily an indication for immediately repeating the pap correlation with the history and clinical exam are dummy the hpv assay was performed in the clinical virology laboratory at mayo clinic this test was developed and its performance characteristics determined by laboratory medicine and pathology mayo clinic rochester mn it has not been cleared or approved by the us food and drug administration gross description a thinprep pap test
HPVQNS	hpv reflex cancelled due to insufficient quantity of testable material
HSIL	high grade squamous intraepithelial lesion
LSIL	low grade squamous intraepithelial lesion
OMN	other malignant neoplasm
UNSAR	unsatisfactory for evaluation

Parser rules map concept codes to parameter values and perform logical checks for validity of the report. The following gives the mapping of the concept codes to classes of interest for cervical cytology type:

Parameter Value	**Concept abbreviation**
unsatisfactory for evaluation	**UNSAR**
abnormal(other than ascus)	**LSIL, HSIL, ASCH, SCC, GECA**
ASCUS	**ASKS**
Negative	**NIL**

Logical validity rules: There is no default value for cervical cytology and its value has to be described in the report. Also the concept for one of the classes can be mentioned in the report as the classes are mutually exclusive.

The following gives the mapping of the concept codes to classes of interest for HPV test:

Parameter Value	**Concept abbreviation**
Positive	**PHPV**
Negative	**NHPV**
Not-performed	**HPVQNS, Default**

If either of the concepts for HPV are not mentioned then the implicit concept class is not performed.

Logical validity rules: The concept for one of the classes can be mentioned in the report as the classes are mutually exclusive.

The following gives the mapping of the concept codes to classes of interest for Endocervical Zone Component:

Parameter Value	**Concept abbreviation**
Inadequate	**NOEC**
Adequate	**Default**

If the CytologyType is NOT "unsatisfactory for evaluation" then the default value is adequate.

For screening reports: The report should contain the word pattern “[\w|\s]+thinprep pap test screen[\w|\s]+”, when the parameter ScreeningReport should be set to value ‘yes’ else the report is not a Screening report. By default the value of this parameter is ‘no’.

B) Detailed overview of the flowchart representing the guideline rulebase

Figure 3 in the paper shows the flowchart abstraction of the guideline rulebase. The figure is expanded on page 6 of this document to show the node numbering that corresponds to the drools rules file "ColoRectalCancerScreening.drl".

The flowchart consisted of 22 nodes (11 leaf nodes) and 20 edges and spanned five different parts of the EMR. A detailed description is as follows.

Node 1: The flow chart begins with a check in the registration section whether the patient is female and alive, to proceed to node 2.

Node 2: The patient provided information (PPI) section is accessed to find whether the response to the question—"Have your menstrual periods changed in anyway or become abnormal to you?" matches the option "No, I have had a hysterectomy". This is to ensure that female patients that have undergone a hysterectomy are not advised to have a routine screening Pap test. The PPI section consists of the patient's response to annually administered questionnaires. If no history of hysterectomy then proceed to node 3

Node 3: Next for patients with no history of hysterectomy, the list of patient documents is then searched to identify "Pap reports" and the latest report is analyzed to determine the cytology type. If the result is "abnormal (other than ASC-US) the patient is referred to Gynecology Clinic (node R1). If type is "ASC-US" (Atypical squamous cells of undetermined significance) or negative then the recent HPV test result is checked (nodes 4 and 5). If report is unsatisfactory for evaluation, the recommendation is to repeat Pap at 3 months (node R11). If report is absent age of patient is checked (node 6).

Node 4: Recent Hpv result is determined by examining the latest pathology report for Pap smear as well as the laboratory system for the Hpv test. The most recent of the test/report from the two sources is access to determine the Hpv test result. If the recent Hpv test is positive, then the patient is referred to Gynecology Clinic (node R2). If the test is negative, Pap should be repeated at 1 year (node R3).

Node 5: If the recent Hpv test is positive, then the previous Hpv test is looked up, both in laboratory system and Pap report like in node 4 (node 7). If it is negative or not-performed then the endocervical zone component is looked up in the Pathology report.

Node 6: If age is more than 21 years then an immediate Pap test is advised, else no recommendation is made

Node 7: If the previous Hpv test is positive, then the previous the patient is referred to Gynecology Clinic (node R4). If it is negative or not-performed or if there is no such report then the recommendation is to repeat the combined Pap/Hpv test at 1 year.

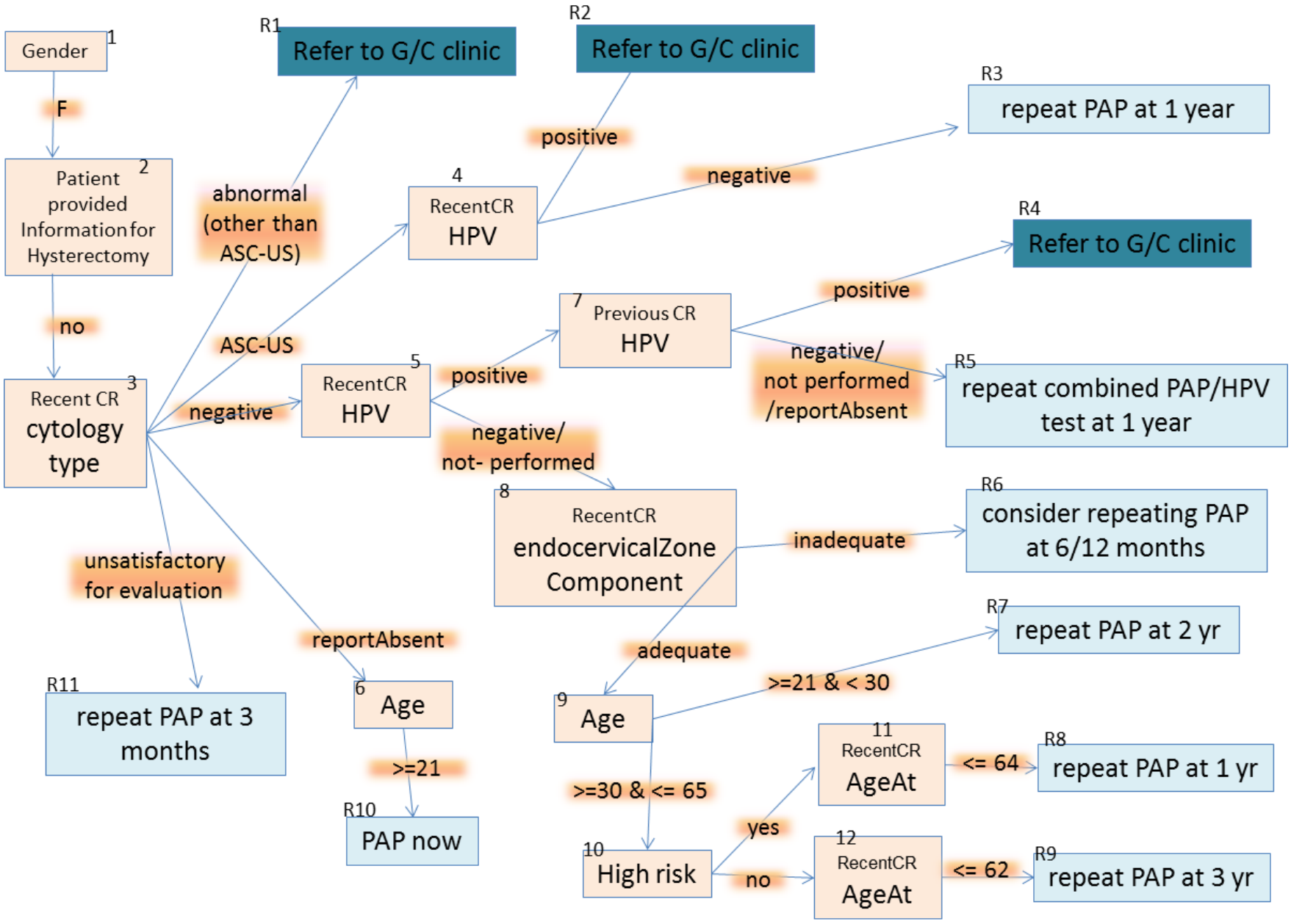
Node 8: If the latest Pap report mentions the finding of an inadequate endocervical zone component then the recommendation is to consider repeating Pap at 6 or 12 months (node R6). If adequate then the age of the patient is considered (node 9).

Node 9: If age is more than or equal to 21 years or less than 30 years then the Pap is to repeated at 2 years from the date of the last Pap test (node R7).

Node 10: The risk status of the patient is determined by checking for whether the conditions following conditions are found in the problem list, and then age at the latest Pap report is checked in nodes 11 and 12, if the risk is present or absent respectively. The conditions are: HIV, cervical dysplasia, Carcinoma cervix, DES exposure as fetus, Hodgkin's disease, multiple myeloma, lymphoma, leukemia, organ transplant and stem cell transplant

Node 11: If age of the patient when the latest Pap was performed is less or equal to 64 years, then the recommendation is to repeat Pap at 1 year after the last Pap report.

Node 12: If age of the patient when the latest Pap was performed is less or equal to 62 years, then the recommendation is to repeat Pap at 3 year after the last Pap report.



Guideline flowchart

C) For the drools implementation of the rules see:

NLPCervicalPathology.drl

CervicalCancerScreening.drl