

Automatic Extraction of Synoptic Data

George Cernile

Artificial Intelligence in Medicine

AIM

Agenda

- Background
- Technology used
- Demonstration

Partial CAP Checklist definitions for 3 Cancer Sites

(Site)

Breast

Specimen Type (Attributes)

Excision
Mastectomy
Other (specify):
Not specified

(Values)

Lymph node sampling

No lymph node sampling
Sentinel lymph node(s) only
Sentinel lymph node with axillary dissection
Axillary dissection

Specimen Size

Greatest dimension: __ cm
*Additional dimensions: _ x _ cm
Cannot be determined

Colorectal

Tumor Site

Cecum
Right (ascending) colon
Hepatic flexure
Transverse colon
Splenic flexure
Left (descending) colon
Sigmoid colon
Rectum
Not specified

Polyp Size

Greatest dimension: __ cm
*Additional dimensions: _ x _ cm
Cannot be determined

Polyp Configuration

Pedunculated with stalk
 Stalk length: __ cm
Pedunculated, no stalk
Sessile
Fragmented

Brain/Spinal Cord

Specimen Type

Open biopsy
Stereotactic needle core biopsy
Subtotal/partial resection
Total resection
Other (specify):
Not specified

Specimen Size

Greatest dimension: __ cm
*Additional dimensions: _ x _ cm

Tumor Site (check all that apply)

Cerebral meninges
Cerebrum (specify lobe[s], if known):
Basal ganglia
Thalamus
Hypothalamus
Suprasellar
Pineal
Cerebellum
Cerebellopontine angle
Ventricle
Brain stem
Spinal cord
Nerve root
Other (specify):

Questions

- How often are checklist elements included in a report, and does this vary by type of report/cancer?
- With what degree of confidence can checklist elements be identified and the associated value read?
- Could a system associate the appropriate set of checklist elements with the cancer that is being reported?
- How much variability is there in the completeness of checklist elements between and within types of cancer?
- What type of lexicon and logic system would be required to encode the checklist elements?
- Will the system be able to determine when data are ambiguous and, if so, how should they be dealt with?

Machine readability criteria

Score	Criteria
1	Checklist element present, value clearly defined
2	Checklist element present, value ambiguous
3	Checklist element not present, value can be inferred
4	Cannot determine value

Examples

Checklist Element	Text	Score	Explanation
Lymph Node Sampling	<i>“The specimen consists of 3 small lymph nodes ranging in size from 0.5 to 1.2 cm. “</i>	2	Lymph nodes are mentioned but not labeled as a sampling, the words “specimen consists” is taken to mean that they are part of the sampling.
Tumor Size	<i>“The specimen is serially sectioned to reveal a grayish white firm tumor with ill defined Margins measuring 2 x 2.2 x 2.8 cm in greatest dimensions.”</i>	2	Not explicit – the measurement might also refer to the margins
Size of Invasive Component	<i>“Invasive Tumour Size 2.0 x 1.7 cm “</i>	1	Invasive Tumor size clearly identified.
Histologic Type	<i>“Sections showed multiple cores of breast tissue showing fibrosis with scattered areas of small ductal epithelial cells arranged in sheets and cords of cells suspicious for invasive carcinoma.”</i>	3	No element identifier present, histologic type can be inferred.

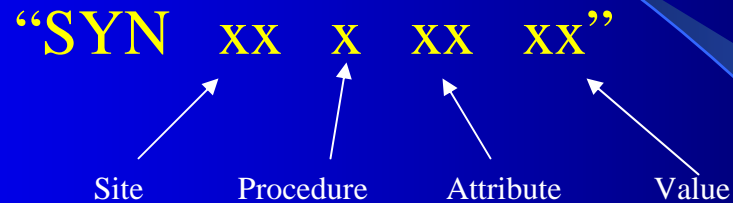
Machine readability data

Checklist Element Name (CE)	No. of Reports Containing CE Identifier	No. of Reports Containing CE Value	Machine Legibility Score Distribution			
			1	2	3**	4
Histologic Grade	40	31	40	15	8	5
Specimen Type	7	55	7		48	13
Lymph Node Sampling	16	39	16	23		
Specimen size	1	63	1		63	4
Laterality	1	66	1		65	2
Tumor Site	2	58	8	1	49	10
Size of Invasive Component	9	29	9		20	
Histologic Type	37	33	33	2	34	
Pathologic Staging	4	19	4		15	
Margins	45	45	45			23

70 pathology reports of breast cancer

**Checklist element not present, value can be inferred

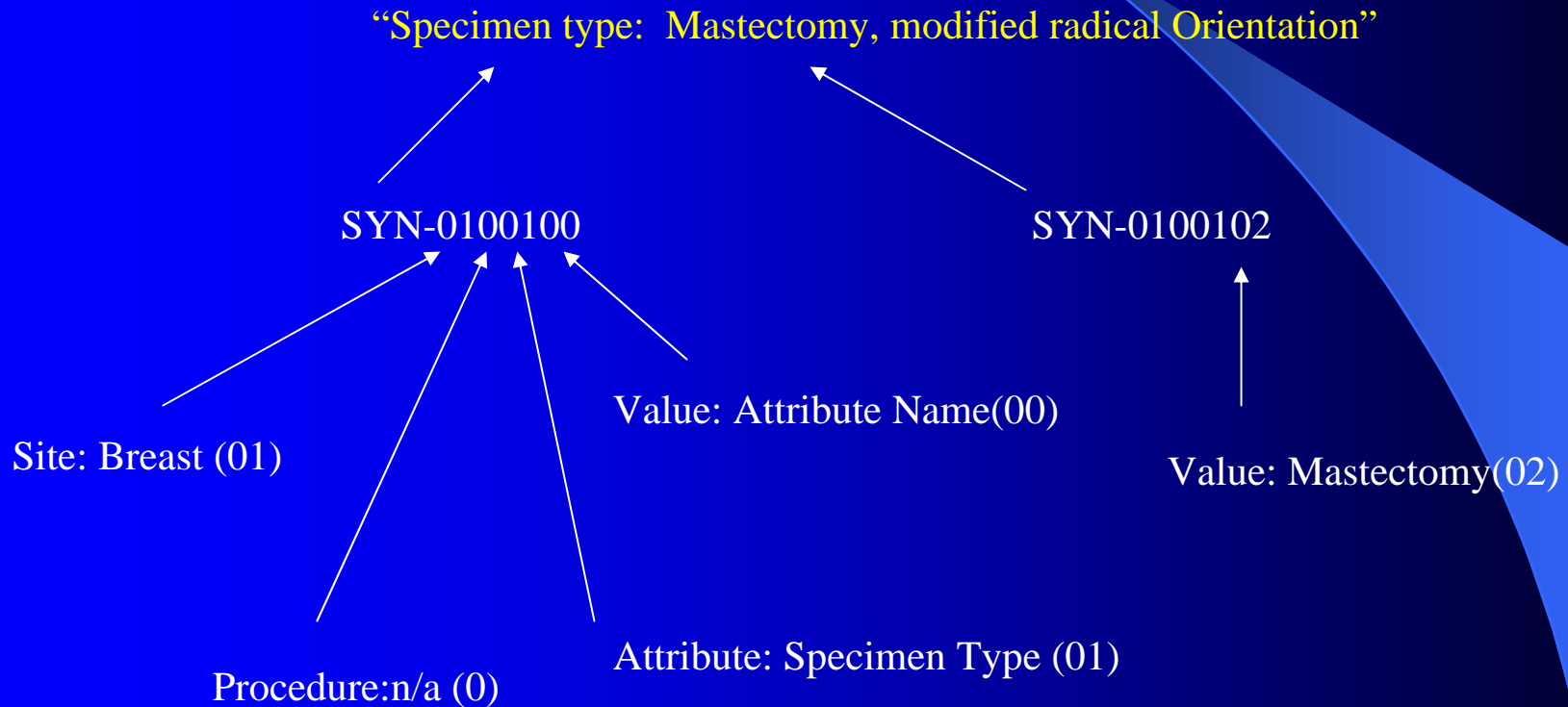
Internal Coding system



e.g. Site: Breast (01)
 Procedure: n/a (0)
 Attribute: Histologic Type (07)
 Value: Non invasive Carcinoma (01)

Code: SYN_0100701

Example from Breast



Synoptic extraction tasks

- **Identify the report type**
 - Lexicons ICD-O-3, etc return concepts found in text
 - Knowledge base infers report type from concepts
- **Extract relevant data elements**
 - Lexicon specific report type returns relevant concepts
 - Numeric extraction algorithm
 - Knowledge base matches values and numbers to elements

Site Identification

- **NHL** M-9590/3 – M-9591/3 , M-9670/3 – M-9699/3, M-9700/3 – M-9719/3
- **Hodgkin's Lymphoma** M-9650/3 – M-9667/3
- **Ewing Sarcoma** M-9260/3
- **Uveal Melanoma** C69.* + M-8772/3 – M-8774/3
- **GI Lymphoma** C16.*, C17.*, C18.*, C19.* C20.*, C21.* + M-9590/3 – M-9591/3 , M-9670/3 – M-9699/3, M-9700/3 – M-9719/3, M- 9650/3 – M-9667/3

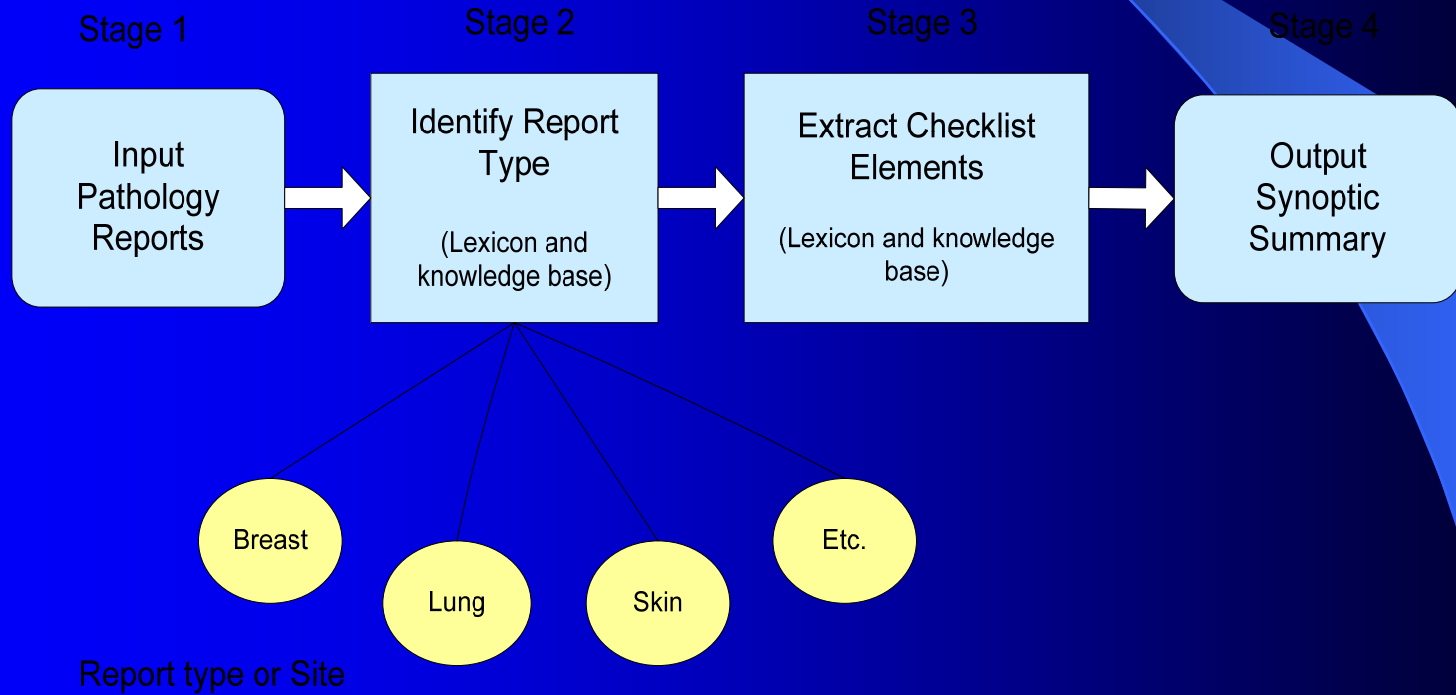
Inference Engine

- Imbedded AI tool for knowledge base
 - CLIPS (developed by NASA)
- Powerful pattern matching and rule based inference engine
- Can express arbitrarily complex knowledge to provide a system with expert level problem solving

Numeric Extraction

- Algorithm designed to extract numeric values
- Knowledge base compares ranges and units
- **Examples**
 - “The tumor measures 1.5 cm.”
 - “breast tissue measuring approximately 5 x 2 x 1 cm”

System block diagram



Live Demonstration

Synoptex Reporting Tool

Start

SYNEX Reports

- 04-SU12014 Colon
- 04-SS6545 Colon
- 04-SU12787 Brain/CNS
- 0 Breast
- A2 Melanoma/Skin
- S05-14959 Lung
- S05-16339 Lung
- S0498 Breast
- S04466 Breast**
- S046658 Urinary Rens
- S03-7517 Colon
- S03-7520 Colon
- S03-20177 Prostate
- S03-24317 Urinary Rens
- S03-20868 Prostate
- S04-5429 Urinary Rens
- S04-2179 Upperaero
- AS03-7616 Skin Carcin
- S03-12451 Colon

Elements found 15/26	Values found 13/26
BREAST Synoptic Report	
*** Specimen Type	Lumpectomy
*** Lymph Node Sampling	Axillary dissection
Lymph Nodes Positive	
Lymph Nodes Examined	
*** Specimen Size (for excisions less th...	2.10x6x1.5 cm
Laterality	Left
Tumor Site	Upper outer quadrant
*** Size of Invasive Component	1.9 cm
*** Histologic Type	Ductal carcinoma in situ (3) Invasive carcinoma (NOS) (8) Invasive ductal carcinoma (9) Apocrine (2)
Grading System	Bloom Richardson
*** Grade	Grade III
*** Tubule Formation	3
*** Nuclear Pleomorphism	3
*** Necrosis	Present
*** Mitotic Count	Greater than 20 mitoses per 10 HPF (score=3)
Pathologic Staging (pTNM)	
HER2 Status	
*** ER Status	Negative
*** PR Status	Negative
*** Margins	Margins uninvolved by invasive carcinoma
Extent of Margin Involvement for Inva...	
Extent of Intraductal Component	
Extent of Margin Involvement for DCIS	
*** Lymphatic/Venous Invasion	Lymphovascular invasion absent
*** Microcalcifications	Microcalcifications present
Additional Pathologic Findings	
Elements found : 15/26	Values found : 13/26

Breast, NOS Lumpectomy, left breast.
Lymph node, NOS Left axillary lymph nodes.

Concordance Code Tissue 1: (1); Concordance Code Tissue 2: (2):

Breast lumpectomy (left) - Invasive ductal carcinoma. Modified Bloom-Richardson Grade 3. Invasive tumour measures 1.9 cm in greatest dimension. Resection margins negative. Invasive tumour measures less than 1 mm from the anterior resection margin. Lymphovascular invasion absent. See synoptic report. Axillary lymph nodes dissection (left) - 13 lymph nodes identified and all negative for malignancy.

Specimen consists of a rounded piece of fibrofatty breast tissue. The specimen is orientated with a short suture marking the superior margin and a long suture marking the lateral margin. No skin ellipse. Resection margin is inked. Serial sectioning reveals a nodular solid mass with relatively ill-defined margins, measuring approximately 2.2 cm. superior/inferiorly, 2 cm. anteroposteriorly, and 2.5 cm. medial/laterally. The nodule closely borders on the anterior margin, and measures approximately 1 cm. from the superior margin, 1.5 cm. from the inferior margin, 1.5 cm. from the lateral margin, 3 cm. from the medial margin, and 0.2 cm. from the posterior margin.

Sections are as follows. Cassette A, lateral margin. Cassette B, medial margin. Cassettes C, D, and E, superior margin. Cassettes F, G, and H, inferior margin. Cassettes I, J and K, anterior margin. Cassettes L and M, posterior margin. Cassettes N, O, and P, sections of nodule. Cassette Q, adjacent breast tissue.

2. Specimen consists of fat tissue measuring 10 x 6 x 1.5 cm. in maximal dimension. Lymph nodes are retrieved and submitted as follows. Cassettes A, B, and C, contain one lymph node divided between these cassettes. Cassette D, contains one lymph node bisected.

Additional lymph nodes submitted in cassettes E and F. Cassette G contains one lymph node bisected.

1,2. Histologic findings are summarized in a synoptic report for invasive breast carcinoma.

Procedure - #1 Breast lumpectomy. #2 Axillary lymph node dissection.
Site - Left breast.
Histologic type - Invasive ductal carcinoma.
Tumour size - Invasive tumour measures 1.9 x 1.4 cm. in greatest dimension (microscopic measurement, block 1 - 0).

Enhanced text extraction

AutoCode Ver. 3.2.7

Viewer Activated

Shows how value was derived

Synoptex Reporting Tool

Start

SYNEX Reports

- 04:SU12014 Colon
- 04:SS8546 Colon
- 04:SU12787 Brain/CNS
- 6 Breast
- A2 Melanoma/Skin
- S05-14959 Lung
- S05-15309 Lung
- S04/98 Breast
- S04/466 Breast
- S04/6858 Urinary Ren
- S03-7517 Colon
- S03-7520 Colon
- S03-20177 Prostate
- S03-24317 Urinary Ren
- S03-20868 Prostate
- S04-5429 Urinary Ren
- S04-2179 Upperaero
- AS03-7618 Skin Carcin
- S03-12451 Colon

Elements found 15/26	Values found 13/26
BREAST Synoptic Report	
*** Specimen Type	Lumpectomy
*** Lymph Node Sampling	Axillary dissection
Lymph Nodes Positive	
Lymph Nodes Examined	
*** Specimen Size (for excisions less th...	2.10x6x1.5 cm
Laterality	Left
Tumor Site	Upper outer quadrant
*** Size of Invasive Component	1.9 cm
*** Histologic Type	Ductal carcinoma in situ (3) Invasive carcinoma (NDS) (8) Invasive ductal carcinoma (9) Apocrine (2)
Grading System	Bloom Richardson
*** Grade	Grade III
*** Tubule Formation	3
*** Nuclear Pleomorphism	3
*** Necrosis	Present
*** Mitotic Count	Greater than 20 mitoses per 10 HPF (score=3)
Pathologic Staging (pTNM)	
HER2 Status	
*** ER Status	Negative
*** PR Status	Negative
*** Margins	Margins uninvolved by invasive carcinoma
Extent of Margin Involvement for Inva...	
Extent of Intraductal Component	
Extent of Margin Involvement for DCIS	
*** Lymphatic/Venous Invasion	Lymphovascular invasion absent
*** Microcalcifications	Microcalcifications present
Additional Pathologic Findings	
Elements found : 15/26	Values found : 13/26

2. **Specimen consists of fat tissue measuring 10 x 6 x 1.5 cm.** in maximal dimension.
Lymph nodes are retrieved and submitted as follows. Cassettes A, B, and C, contain one lymph node divided between these cassettes. Cassette D, contains one lymph node bisected.
Additional lymph nodes submitted in cassettes E and F. Cassette G contains one lymph node bisected.

1,2. Histologic findings are summarized in a synoptic report for invasive breast carcinoma.

Procedure - #1 Breast lumpectomy. #2 Axillary lymph node dissection.
Site - Left breast.
Histologic type - Invasive ductal carcinoma.
Tumour size - Invasive tumour measures 1.9 x 1.4 cm. in greatest dimension (microscopic measurement, block 1 - 0).
Resection margins - Negative. Invasive tumour less than 1 mm. from anterior margin (blocks 1 - I, J, K).
Tumour border - Infiltrating.
Tubule formation - Grade 3.
Nuclear Grade - Grade 3.
Mitotic Grade - Grade 3 (field diameter = 0.625 mm).
Overall Grade - Grade 3.
Necrosis - Present, multifocal.
Lymphovascular invasion - Absent.
Perineural invasion - Absent.
Intraductal component - Present. Solid pattern, nuclear grade 3. Comedo pattern, focal.
Extensive DCIS - Absent.
Lymphocytic response - Moderate.
Nipple involvement - Not applicable.
Skeletal muscle involvement - Not applicable.
Skin involvement - Not applicable.
Other lesions - Cancerization of lobules. Fibrocystic change consisting of stromal sclerosis, cyst formation, and apocrine change.
Microcalcifications - Stromal microcalcifications present within fibrocystic change.
Axillary lymph nodes - 13 lymph nodes identified and all negative for malignancy.
ER/PR status - ER and PR immunohistochemistry performed in Hamilton. See Henderson Hospital report.

Enhanced text extraction

AutoCode Ver. 3.2.7

Viewer Activated

Uses for Synoptex

- API for embedded synoptic extraction capability
- Add-on to E-path for automatic synoptic report generation
- Quality control tool for reporting institutions
- Automated data extraction to populate database for clinical trial matching system
- Other uses where synoptic summaries are needed (ICU report scanning)