Expression and cut parser for CMS event data

Luca Lista, INFN Sezione di Napoli, **Christopher D. Jones**, FNAL,
Giovanni Petrucciani, Scuola Normale Superiore di Pisa (SNS) and INFN Sezione di Pisa

We present a parser to evaluate expressions and boolean selections that is applied on CMS event data for event filtering and analysis purposes. The parser is based on boost spirit grammar definition, and uses Reflex dictionary for class introspections. The parser allows a natural definition of expressions and cuts in users configuration, and provides good run-time performances compared to other existing parsers.

Expression/cut

```
pt>15 && abs(eta)<2
```
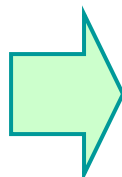
Object model
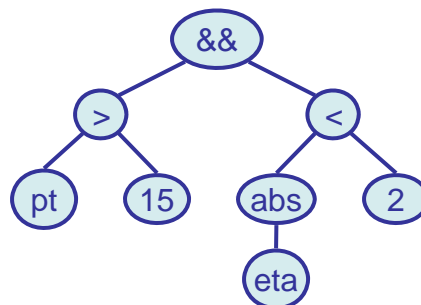
Cut and expression parsing:
is a core part of the Analysis Tools providing a flexible user configurability

Grammar definition:
written in C++ with Boost Spirit library

Support for many math functions, operators, logical comparator and parentheses

Object method mapping:
introspection using Reflex dictionaries

Natural variable to method mapping:

```
pt           → object.pt()
daughter(0) → object.daughter(0)
```

Accepted arguments: all integer and floating Point types, std::string (via **"string"**) and enumerators (via **'enumName'**)

Nested method calls:
```
track.pt    → object.track().pt()
```
Automatic check for pointers and persistent reference types supported:
```
object.track()->pt()
```

Use in C++

```
StringCutObjectSelector<reco::Track>
  select("pt>15.0 && abs(eta)<2");

StringObjectFunction<reco::Track>
  f("px^2 + py^2");

reco::Track trk = …;

bool pass = select(trk);
double ptSquare = f(trk);
```

Usage in CMS selectors/filters python config.

```
zCandidates = cms.EDFilter(
   "CandSelector",
   src = cms.InputTag("dimuons"),
   cut = cms.string("min(daughter(0).pt,daughter(1).pt)>20 &&"
   "fabs(daughter(0).eta)<2 && fabs(daughter(0).eta)<2 &&"
   "daughter(0).isGlobalMuon=1 && daughter(1).isGlobalMuon=1"
)
```

Generic histogramming modules

```
zPlots = cms.EDAnalyzer(
  "CandHistoAnalyzer",
  src = cms.InputTag("zCandidates")
  histograms = cms.VPSet(
    cms.PSet(
      min = cms.untracked.double(0.0),
      max = cms.untracked.double(200.0),
      nbins = cms.untracked.int32(200),
      name = cms.untracked.string("zMass"),
      description = cms.untracked.string("Z mass [GeV/c^{2}]"),
      plotquantity = cms.untracked.string("mass")
    ),
    cms.PSet(
      min = cms.untracked.double(0.0),
      max = cms.untracked.double(200.0),
      nbins = cms.untracked.int32(200),
      name = cms.untracked.string("mu1Pt"),
      description =
        cms.untracked.string("Highest muon p_{t} [GeV/c]"),
      plotquantity =
        cms.untracked.string("max(daughter(0).pt,daughter(1).pt)")
    ),
  )
)
```

Expressions

Used in CMS Physics Analysis Toolkit and Event Display (Fireworks) for interactive/batch Object selection

Performances

- Parsing and method look-up done once at object construction time
- Up to ~50 times faster than ROOT TFormula in some cases